7

ENCICLOPEDIA IBEROAMERICANA DE FILOSOFÍA

Lógica

Edición de
Carlos E. Alchourrón
José M. Méndez
Raúl Orayen

Editorial Trotta
Consejo Superior de Investigaciones Científicas

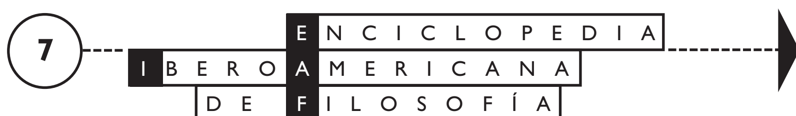


Lógica

Lógica

Edición de Carlos E. Alchourrón

Editorial Trotta
Consejo Superior de Investigaciones Científicas



Primera edición: 1995
Primera reimpresión: 2005

Ferraz, 55. 28008 Madrid
Teléfono: 91 543 03 61
Fax: 91 543 14 88
E-mail: editorial@trotta.es
http://www.trotta.es

Departamento de Publicaciones
Vitruvio, 8. 28006 Madrid
Teléfono: 91 561 62 51
Fax: 91 561 48 51
E-mail: publ@orgc.csic.es

Diseño
Joaquín Gallego

ISBN: 978-84-87699-48-1 (Obra completa)
ISBN (edición digital pdf): 978-84-9879-392-5 (vol. 7)

La Enciclopedia IberoAmericana de Filosofía es un proyecto de investigación y edición, puesto en marcha por el Instituto de Filosofía del Consejo Superior de Investigaciones Científicas (Madrid), el Instituto de Investigaciones Filosóficas de la Universidad Nacional Autónoma de México y del Centro de Investigaciones Filosóficas (Buenos Aires), y realizado por filósofos que tienen al español por instrumento lingüístico.

Existe una pujante y emprendedora comunidad filosófica hispanoparlante que carece, sin embargo, de una obra común que orqueste su plural riqueza y contribuya a su desarrollo. No se pretende aquí una enciclopedia de filosofía española sino articular la contribución de la comunidad hispanoparlante a la filosofía, sea mediante el desarrollo cualificado de temas filosóficos universales, sea desentrañando la modalidad de la recepción de esos temas filosóficos en nuestro ámbito lingüístico.

La voluntad del equipo responsable de integrar a todas las comunidades filosóficas de nuestra área lingüística, buscando no sólo la interdisciplinariedad sino también la internacionalidad en el tratamiento de los temas, nos ha llevado a un modelo específico de obra colectiva. No se trata de un diccionario de conceptos filosóficos ni de una enciclopedia ordenada alfabéticamente sino de una enciclopedia de temas monográficos selectos. La monografía temática permite un estudio diversificado, como diverso es el mundo de los filósofos que escriben en español.

La Enciclopedia IberoAmericana de Filosofía es el resultado editorial de un Proyecto de Investigación financiado por la Comisión Interministerial de Ciencia y Tecnología y por la Dirección General de Investigación Científica y Técnica del Ministerio de Educación y Ciencia. Cuenta también con la ayuda de la Consejería de Educación y Cultura de la Comunidad Autónoma de Madrid.

CONTENIDO

CONCEPCIONES DE LA LÓGICA[1]

Carlos E. Alchourrón

I. INTRODUCCIÓN

So long as the sciences are imperfect, the definitions must partake of their imperfection; and if the former are progressive, the latter ought to be so too (John Stuart Mill).

Los textos tradicionales de lógica usualmente comenzaban con una caracterización de la lógica, seguida de una detallada comparación de su contenido y enfoque con los de otras disciplinas estrechamente vinculadas a ella. Esta costumbre en gran medida se ha perdido. En los textos contemporáneos es frecuente encontrar sólo unas breves consideraciones referidas a la definición de la lógica y muy pocas comparaciones, en muchos casos totalmente ausentes, con la temática de otras disciplinas afines. Esta evolución en cuanto a la disminución de la extensión dedicada a la definición de la disciplina y su comparación con otras es un rasgo que acompaña al enriquecimiento intrínseco de toda ciencia. Cuanto más abundante es el material a exponer en una ciencia menos es el espacio que se reserva a la definición de su área temática y al deslinde con otras ciencias. Estos últimos objetivos pasan a integrar, entonces, los temas de la filosofía de la ciencia en cuestión. La relativa autonomía que en cada ciencia se produce respecto de su correspondiente filosofía como consecuencia de su propio desarrollo puede interpretarse como un síntoma de madurez, en la medida que por un lado permite al científico continuar con su tarea sin verse embarcado en complicadas cuestiones

1. Para una presentación diferente de la idea central de este ensayo ver C.E. Alchourrón y A. A. Martino, «Lógica sin verdad»: *Theoria,* 3 (1987/88) 459-464, San Sebastián.

Quiero agradecer las importantes sugerencias y comentarios de David Makinson y Thomas Moro Simpson que permitieron mejorar considerablemente el contenido y la estructura de este escrito.

filosóficas, y por otro, permite al filósofo profundizar sus problemas específicos apoyándose en los resultados de la ciencia. Sin embargo, no son pocos los momentos en que el desarrollo mismo de una ciencia depende de una adecuada reflexión filosófica sobre el área temática de la disciplina. Tal es el caso de la lógica en su último siglo de vida. Caracterizaciones aceptadas durante siglos fueron desplazadas por otras como resultado del desarrollo mismo de la lógica. Sin embargo, este cambio ocurrió sin abandonar el núcleo central que define el área temática de la disciplina.

El propósito de este ensayo, ubicado al comienzo del volumen sobre lógica en una enciclopedia general de filosofía, es tomar en cuenta, aunque sea brevemente, alguno de los tópicos filosóficos vinculados al deslinde de la lógica.

No es tarea fácil la de dar una definición del área temática de una disciplina, cualquiera que ella sea. Esto corrientemente se debe a que por un lado el origen histórico de las distintas disciplinas es fijado, más o menos arbitrariamente por los historiadores, destacando los trabajos de alguno o algunos autores representativos como las obras iniciales de la disciplina, pero con clara consciencia de que ellos fueron precedidos por observaciones y descubrimientos en el área cuyos autores se desconocen o tienen menos importancia. Además, es frecuente que el desarrollo histórico del cuerpo teórico de cada disciplina haya sido gradual y acumulativo, tal como ha ocurrido, por ejemplo, con el contenido teórico de ciencias como la astronomía, la física o la matemática que son el resultado de adiciones y rectificaciones acumuladas a lo largo de siglos por un sinnúmero de autores, de importancia diversa, y que ha llevado, en muchos casos, a ampliaciones y cambios más o menos significativos de la temática históricamente inicial.

La historia de la lógica es, en la dirección apuntada, radicalmente distinta. Su origen histórico tiene fecha cierta. La lógica es una teoría que se inicia en los libros del *Organon* de Aristóteles. La teoría del Silogismo Categórico, contenida sustancialmente en los *Primeros Analíticos,* es y sigue siendo el paradigma para identificar la temática de la lógica, aunque no, por cierto, su contenido teórico, que se ha incrementado enormemente y se ha modificado en rasgos importantes. Además, la historia de la lógica está signada por discontinuidades tan marcadas que se hace difícil hallar paralelos en la historia de otras ciencias. Esta curiosidad en la historia de la lógica está enfatizada, aunque errónea y exageradamente, por Kant en el prólogo a la segunda edición de su *Crítica de la Razón Pura* cuando presenta a la lógica como una ciencia que nació perfecta y completa en manos de su creador: Aristóteles. Aunque Kant se equivocó en esta apreciación histórica, ya que desde Aristóteles hasta los días de Kant la teoría lógica fue objeto de múltiples, y en algunos casos, sustanciales modificaciones, hay mucho de cierto en la imagen kantiana de la historia de la lógica, ya que, sin exagerar demasiado, puede afirmarse que el cuerpo central de la teoría lógica contemporánea surgió en las postrimerías del siglo diecinueve en las obras de Frege, sin que pueda

señalarse en ellas influencia alguna ni continuidad con el enfoque teórico aristotélico, y en cierto sentido, oponiéndose a éste. Sin embargo, es dable señalar un tema central que, fijado en la obra de Aristóteles, permanece idéntico hasta nuestros días de modo que hace factible una definición general del área temática de la lógica.

II. EL ENFOQUE PSICOLÓGICO

En una primera aproximación la lógica deductiva (ya que ése es nuestro tema) puede describirse como la teoría de los razonamientos (deductivos). Esta caracterización, además de excesivamente imprecisa, tiene un inadecuado cariz psicológico que ha tenido, y que en ciertos enfoques de la lógica (el que recibe contemporáneamente en muchos trabajos de inteligencia artificial) continúa teniendo una persistente influencia que entorpece la identificación de la lógica.

Una definición muy corriente en las obras escritas antes de nuestro siglo identificaba la lógica con la ciencia y/o el arte del pensamiento. Tal es el caso de la muy influyente obra de Antoine Arnauld y Pierre Nicole (1662) *La Logique ou l'Art de penser* conocida como *La Lógica de Port Royal*. En este sentido es también significativo el título *An Investigation of the Laws of Thought on which are founded the Mathematical Theories of Logic and Probabilities (Laws of Thought)* con que George Boole designó, en 1854, a uno de los libros más influyentes en la lógica contemporánea.

De este modo, la vinculación de la lógica con la psicología fue, desde la perspectiva de muchos y muy representativos autores, tan estrecha que identificaban a la lógica como describiendo, y a veces prescribiendo, ciertos procesos psicológicos (razonamientos, argumentaciones) en que estaban involucrados estados psicológicos de los individuos (juicios, creencias, conocimientos).

Es realmente asombroso que una caracterización, en cierto modo, tan errada como la que ofrece la definición psicologista de la lógica tuviera un consenso tan amplio y duradero, que fuera necesario para descartarla la enorme tarea y dedicación que a fines del siglo pasado y principios del actual pusieron Frege y Husserl en la lucha antipsicologista en cuanto a la definición de la lógica. En el caso de Frege sus argumentos antipsicologistas estuvieron en gran medida focalizados en la definición psicologista de la matemática, y más en particular de la aritmética, pero no hay que olvidar que para Frege la matemática no es más que el capítulo más avanzado de la lógica.

Sería realmente insensato intentar justificar las leyes de cualquier teoría lógica apoyándose en las propiedades que pudieran descubrirse observando los procesos psicológicos efectivos de argumentación que los hombres realizan a diario. El resultado de tal investigación, de naturaleza claramente empírica y contingente, que seguramente exhibiría características

muy distintas frente a individuos de grupos humanos heterogéneos, sería impotente para dar cuenta del carácter necesario y a priori de las leyes lógicas, quizás sólo seriamente cuestionado por John Stuart Mill. Sin embargo, la larga tradición psicologista en la definición de la lógica ha tenido una enorme influencia, si no en el desarrollo de la teoría lógica, en el vocabulario usado en la formulación de las teorías lógicas. En efecto, verbos como inferir, argumentar, deducir, etc., designan indudablemente procesos psicológicos que los hombres realizan con frecuencia. A su vez, los sustantivos correspondientes: inferencia, argumento, deducción, etc., y a pesar de su clara ambigüedad proceso-producto, conservan en su designación la connotación psicológica de los verbos asociados. Es más, si bien en lógica se ha acuñado la expresión «premisa(s)» para indicar los puntos de partida de una inferencia, se sigue usando la expresión «conclusión», con su clara connotación de punto final de un proceso (en este caso psicológico), para referirse a lo que se pretende estar justificado por las premisas en un esquema inferencial.

Si bien, en principio, carece de justificación la definición de la lógica como ciencia teórico-descriptiva del pensamiento, es en cambio más plausible aquella definición (vinculada a la idea de «arte del pensar») que caracteriza a la lógica como una disciplina normativa destinada a prescribir cómo se debe pensar (argumentar, inferir) para hacerlo correctamente. Desde esta perspectiva el objeto de la lógica sería, no ya describir cómo los hombres efectivamente argumentan, sino efectuar una suerte de control de calidad con relación al producto de la actividad argumentativa, codificando los esquemas argumentativos que logran, distinguiéndolos de los que no logran, la finalidad implícita en la actividad argumentativa. Este enfoque presupone que se explicite la o las finalidades que la lógica toma en cuenta en su control de calidad de los procesos argumentativos, ya que, como toda actividad, los procesos de argumentar pueden llevarse a cabo guiados por las finalidades más diversas. Así, una argumentación realizada con el propósito de persuadir a alguien, será buena o mala si de hecho se logra con ella convencer a la persona a la que está dirigida. Esta finalidad persuasiva que puede ser importante para juzgar sobre el valor retórico de una argumentación no es por cierto la finalidad contemplada en la lógica. En el enfoque que estamos considerando se asume que la finalidad (por lo menos la finalidad que la lógica tomará en cuenta) de una argumentación será preservar en la conclusión la verdad de las premisas. El objetivo de la lógica sería entonces encontrar criterios que aseguren la verdad de la conclusión para el caso en que las premisas sean verdaderas.

III. EL ENFOQUE SEMÁNTICO

Siguiendo el camino anterior, que transita por las huellas de la tradición del «arte de pensar», puede llegarse a una de las caracterizaciones más

representativas de la visión actual frente a la cuestión de la identificación temática de la lógica: el enfoque semántico de la noción de consecuencia. Por una inferencia se entenderá desde ahora un conjunto de enunciados, de un lenguaje previamente especificado, en el que la verdad de uno de ellos (la conclusión de la inferencia) se pretende justificar en la verdad de los otros (las premisas de la inferencia). La inferencia será buena (válida) cuando la conclusión sea consecuencia necesaria de las premisas, o lo que es lo mismo, cuando las premisas impliquen lógicamente la conclusión. Esta idea puede resumirse en cualquiera de las siguientes dos definiciones intuitivas que servirán como punto de partida para lograr otras técnicamente más precisas.

(Def. 1.0) Un enunciado C es consecuencia del conjunto de premisas $P_1...P_n$ si y sólo si es imposible que las premisas $P_1...P_n$ sean todas verdaderas y la conclusión C no lo sea, o equivalentemente:

(Def. 1.1) Un enunciado C es consecuencia del conjunto de premisas $P_1...P_n$ si y sólo si es necesario que si todas las premisas son verdaderas la conclusión también lo sea.

Es claro que cuando se cumple la condición expuesta, la verdad de las premisas justifica la verdad de la conclusión, es decir, se cumple con la finalidad, considerada por la lógica en todo proceso (psicológico) argumentativo, de preservar en la conclusión la verdad de las premisas.

En la noción de consecuencia que estamos comentando hay dos tipos de expresiones que requieren especiales aclaraciones: por un lado están las nociones modales de necesidad e imposibilidad, y por otro las de verdad y falsedad. En este momento nuestro propósito es presentar esquemáticamente el enfoque de la lógica que deriva de los trabajos de A. Tarski sobre el concepto de verdad (Tarski, 1935) y el concepto de consecuencia lógica (Tarski, 1936).

Comencemos con las nociones de verdad y falsedad. En el enfoque tarskiano verdad y falsedad son calificaciones hechas en el metalenguaje que versa acerca de las expresiones de un lenguaje (objeto) $\mathbb{L}$ a los enunciados $\mathbb{L}$. En este enfoque los «portadores de la verdad» son expresiones lingüísticas (los enunciados del lenguaje objeto $\mathbb{L}$). No son estados psicológicos ni el significado (proposiciones) de tales expresiones lingüísticas. Sin embargo, para que pueda atribuirse un valor de verdad (verdad o falsedad) a un enunciado éste tiene que ser un enunciado significativo y esto supone que el lenguaje tiene que estar interpretado a través de alguna correlación (explicitada en la parte semántica del metalenguaje) de algunas de sus expresiones con las entidades de la realidad acerca de las cuales versa el lenguaje objeto $\mathbb{L}$.

Por razones que por ahora no vamos a analizar, Tarski considera que por sus peculiares características ninguno de los lenguajes naturales (español, inglés, portugués, alemán, etc.) admite una noción de interpretación con el grado de precisión que se requiere para dar una explicación coherente y satisfactoria de la noción de verdad (y de falsedad). Por esta razón, su construcción está referida siempre a un lenguaje artificialmente

creado, en donde no existen las imprecisiones sintácticas y semánticas de los lenguajes naturales. En ésto la obra de Tarski está signada por uno de los rasgos distintivos de la tarea lógica en la última centuria: crear y estudiar lenguajes artificiales con el propósito de reconstruir en ellos algunas propiedades (no todas) de las expresiones de los lenguajes naturales. Además, en ésto la lógica no hace más que seguir el camino de las ciencias más avanzadas, en efecto, cuando ellas tienen que dar cuenta de una realidad compleja comienzan por construir un modelo simplificado en el que sólo se representan los aspectos que interesan, dejando fuera todo lo demás.

Supongamos un lenguaje artificial $\mathbb{L}$ con la siguiente super simple estructura sintáctica. El vocabulario de $\mathbb{L}$ está integrado por los signos de las siguientes cuatro categorías sintácticas:

Nombres: $a_1 \ldots a_n \ldots$

Predicados (monádicos): $P_1 \ldots P_n$

Signos lógicos:

$\neg$ (Negación)
$\wedge$ (Conjunción)
$\vee$ (Disyunción incluyente)
$\supset$ (Condicional material)

Signos de puntuación: «(»y«)» (paréntesis izquierdo y derecho).

Los enunciados de $\mathbb{L}$ serán las secuencias de signos de $\mathbb{L}$ (expresiones de $\mathbb{L}$) que satisfacen alguna de las siguientes cláusulas (reglas de formación [de enunciados] de $\mathbb{L}$):

1. Enunciados atómicos: si P es un predicado de $\mathbb{L}$ y *a* es un nombre de $\mathbb{L}$, entonces P*a* (P seguido de *a*) es un enunciado (atómico) de $\mathbb{L}$.
2. Enunciados moleculares: si A y B son enunciados de $\mathbb{L}$, entonces $\neg$ A, (A $\wedge$ B), (A $\vee$ B) y (A $\supset$ B) son enunciados (moleculares de $\mathbb{L}$).

Las reglas anteriores 1 y 2, que pertenecen al metalenguaje sintáctico de $\mathbb{L}$, especifican cuáles de las expresiones de $\mathbb{L}$ son sus enunciados, pero nada dicen acerca del significado (en el sentido de referencia a la realidad) de ninguna de las expresiones de nuestro lenguaje objeto. Para este último propósito supongamos que contamos en nuestro metalenguaje con funciones del tipo $| \quad |_i$[2] (funciones de interpretación) cada una de las cuales correlaciona cada uno de los nombres del lenguaje con un objeto y sólo uno de la realidad y cada uno de los predicados de $\mathbb{L}$ con una clase de objetos de la realidad y sólo una. Si *a* es un nombre, $| a |_i$ es el objeto nombrado por *a* en la interpretación $| \quad |_i$; si P es un predicado, $| P |_i$

2. Lo indicado en el texto es instrumental suficiente para la semántica de una lógica proposicional. Cuando se trata de una lógica (como la de la cuantificación) en la que figuran los cuantificadores estándar es necesario asociar a cada función de interpretación $| \quad |_i$ un conjunto no vacío de objetos D_i (llamado dominio de la interpretación $| \quad |_i$) que fija el ámbito de las entidades tomadas en cuenta por los cuantificadores.

es el conjunto de los objetos denotados por el predicado (la extensión de P) en la interpretación $|\quad|_i$. Con estos elementos estamos en condiciones de especificar las condiciones en que son verdaderos los enunciados atómicos en cada una de las interpretaciones de $\mathbb{L}$ (con relación a cada una de las funciones interpretativas $|\quad|_i$ del metalenguaje de $\mathbb{L}$). La cláusula que cumple con tal tarea es:

(i) Un enunciado atómico P*a* es verdadero en la interpretación $|\quad|_i$ si y sólo si $|a|_i \in |P|_i$ (el objeto asignado al nombre «*a*» en la interpretación $|\quad|_i$ es uno de los elementos de la clase asignada al predicado «P» por esa misma interpretación).

Si además suponemos que cada una de las funciones de interpretación cumple las siguientes cláusulas definitorias de las condiciones de verdad (definición contextual del significado de los signos lógicos de $\mathbb{L}$) de los enunciados moleculares, entonces habremos especificado las condiciones en que son verdaderos o falsos todos y cada uno de los enunciados de $\mathbb{L}$ en cada una de sus interpretaciones.

(ii) $\neg$ A es verdad en $|\quad|_i$ si y sólo si A no es verdad en $|\quad|_i$.

(iii) (A $\wedge$ B) es verdad en $|\quad|_i$ si y sólo si tanto A como B son verdad en $|\quad|_i$.

(iv) (A $\vee$ B) es verdad en $|\quad|_i$ si y sólo si A, B o ambas son verdad en $|\quad|_i$.

(v) (A $\supset$ B) es verdad en $|\quad|_i$ si y sólo si A no es verdad en $|\quad|_i$ o B es verdad en $|\quad|_i$.

o

Lo anterior será todo lo que diremos por el momento en cuanto a la noción de verdad requerida para explicar la caracterización anterior de consecuencia lógica en la tradición semántica tarskiana. Sin embargo, quedan por aclarar las nociones modales que figuran en las anteriores definiciones. Tanto la noción de necesidad como la de imposibilidad son reconstruidas, en este enfoque, como generalizciones a partir de las funciones de interpretación $|\quad|_i$ admisibles para $\mathbb{L}$, es decir, las que cumplen las condiciones estipuladas en las cláusulas anteriores de (i) hasta (v).

De este modo se dirá:

(Def. 2.0) Un enunciado A de $\mathbb{L}$ es consecuencia (semántica) del conjunto de enunciados α de $\mathbb{L}$ (premisas), que abreviaremos: $\alpha \vDash$ A, si y sólo si no hay una interpretación $|\quad|_i$ admisible de $\mathbb{L}$ (imposibilidad) en la que todos los enunciados de α son verdaderos y en la que A no lo es.

O, lo que es equivalente:

(Def. 2.1) Un enunciado A de $\mathbb{L}$ es consecuencia (semántica) del conjunto de enunciados α de $\mathbb{L}$ (premisas), $\alpha \vDash$ A, si y sólo si A es verdadera en toda interpretación admisible $|\quad|_i$ de $\mathbb{L}$ (necesidad) en la que son verdaderos todos los enunciados de α.

En resumen, la transición de las definiciones (Def. 1.0) y (Def. 1.1) a las definiciones (Def. 2.0) y (Def. 2.1) está signada por varias características destinadas a obtener resultados más precisos:

(A) En las últimas definiciones la noción de consecuencia está referida a un lenguaje artificial $\mathbb{L}$ (en el que se aspira a reflejar importantes rasgos de los lenguajes corrientes), que cuenta con una estructura sintáctica mucho más perfilada y simple que la de los lenguajes naturales.

(B) La noción de interpretación (las funciones $|\quad|_i$ referidas a $\mathbb{L}$) de las expresiones de $\mathbb{L}$ carece de las ambigüedades y vaguedades que adolecen las expresiones correlativas de los lenguajes corrientes (con este alcance suele decirse que las definiciones (Def. 2.0) y (Def. 2.1) son «reconstrucciones racionales» de (Def. 1.0) y (Def. 1.1)).

(C) La noción de verdad (y falsedad) usada en las últimas definiciones depende: 1) de las funciones de interpretación $|\quad|_i$ (en el sentido en que un mismo enunciado A de $\mathbb{L}$ puede ser verdadero en una interpretación $|\quad|_i$ y falso en otra $|\quad|_j$) y 2) de las cláusulas que la gobiernan (en nuestro ejemplo, las cláusulas que van de (i) a (v)).

(D) Además, la noción de verdad de Tarski pretende reconstruir, fundamentalmente a través de la satisfacción de la condition (i), la noción filosófica de origen aristotélico, de «verdad como correspondencia».

(E) Las nociones modales intuitivas de necesidad e imposibilidad lógica son reconstruidas por medio de cuantificaciones universales sobre la totalidad de funciones de interpretación $|\quad|_i$ admisibles para el lenguaje $\mathbb{L}$. La referencia a la imposibilidad (lógica) de (Def. 1.0) es reemplazada en (Def. 2.0) por «no hay interpretación $|\quad|_i$ admisible» y la referencia a la necesidad (lógica) de (Def. 1.1) es reemplazada en (Def. 2.1) por «para toda interpretación $|\quad|_i$ admisible».

La presente relación $\vDash$ de consecuencia semántica cumple con las siguientes propiedades (de fácil verificación a partir de las definiciones anteriores):

($\vDash$1) Reflexividad generalizada: $\alpha \vDash A$ si $A \in \alpha$.
($\vDash$2) Corte: si $\alpha \vDash B$ y $\alpha \cup \{B\} \vDash A$ entonces $\alpha \vDash A$.
($\vDash$3) Monotonía: si $\alpha \vDash A$ entonces $\alpha \cup \beta \vDash A$.

Ninguna de estas propiedades depende de las específicas cláusulas que van de (i) a (v) a que se encuentra sujeta en el ejemplo anterior la noción de verdad. Cada una de esas propiedades puede probarse recurriendo sólo a (Def. 2.0) o a su equivalente (Def. 2.1).

La noción de «verdad lógica» queda como el caso límite de la relación de consecuencia semántica en el que el conjunto de premisas es vacío. Un enunciado A expresa una verdad lógica: $\vDash A$ (abreviatura de $\emptyset \vDash A$) si y sólo si A es verdadero en todas las interpretaciones $|\quad|_i$ admisibles de $\mathbb{L}$. A su vez, si el lenguaje cuenta con el signo $\supset$ de implicación material que satisface la cláusula (v), la noción de consecuencia semán-

tica es caracterizable a partir de la noción de verdad lógica. En efecto, es fácil verificar el siguiente enunciado metalingüístico de correlación entre las nociones de consecuencia semántica y de verdad lógica, para el caso en que el conjunto de premisas α no sea vacío (que es el caso diferencial entre ambas nociones) y la relación $\vDash$ sea *compacta*[3]:

(P-5) $\alpha \vDash A$ si y sólo si hay en α un conjunto finito de enunciados $A_1 \ldots A_n$ tal que $\vDash ((A_1 \supset (A_2 \supset \ldots (A_n \supset A) \ldots)$.

Con estos elementos se constituye el, actualmente más dominante, enfoque en cuanto a la identificación temática de la lógica. Esta propuesta podría llamarse el paradigma Tarski-Carnap, ya que si bien se debe a Tarski tanto la definición semántica de verdad, como la caracterización de la noción semántica de consecuencia e, indirectamente, la noción de verdad lógica, fue Carnap quien enfatizó la identificación de la lógica como la teorización de esa noción de consecuencia semántica y de verdad lógica. De hecho Carnap, siguiendo ideas fundamentalmente originadas en el formalismo de Hilbert, había sostenido, hasta el momento de la publicación de trabajos de Tarski, en su obra *Sintaxis Lógica del Lenguaje* (Carnap, 1937) la identificación de la lógica como la teorización de una noción puramente sintáctica de consecuencia.

Es más, generalizando esta tesis acerca de la naturaleza sintáctica de la lógica, Carnap (1937) sostuvo que la filosofía no es más que la sintaxis lógica del lenguaje de la ciencia. Luego de las publicaciones de Tarski cambió su paradigma adoptando, en primer término, la tripartición de la metateoría de todo lenguaje en: sintaxis (parte de la metateoría que considera solamente las propiedades y relaciones de los signos de un lenguaje $\mathbb{L}$ con independencia de toda interpretación del lenguaje), semántica (parte de la metateoría en la que se consideran las propiedades y relaciones entre los signos de un lenguaje $\mathbb{L}$ que dependen de las correlaciones entre las expresiones del lenguaje y la realidad establecidas por las funciones de interpretación $|\quad|_i$) y pragmática (parte de la metateoría en la que se consideran las reglas de uso de los signos lingüísticos adoptados por un hablante o una comunidad de hablantes del lenguaje). Semiótica (teoría general de los signos) es el nombre genérico que cubre a la sintaxis, la semántica y la pragmática.

Desde la perspectiva anterior Carnap (1942) reformuló sus tesis de la siguiente manera: la filosofía no es más que la semiótica del lenguaje de la ciencia y la lógica es, como en la caracterización tarskiana, la teoría de la relación de consecuencia semántica y de la noción semántica de verdad lógica.

La influencia de las obras de Tarski fue enorme, y si bien ahora nos estamos refiriendo principalmente a la incidencia que tuvieron en la meta-

3. Una relación de consecuencia $\vDash$ se dice que es compacta si y sólo si cuando un enunciado A es consecuencia de un conjunto α ($\alpha \vDash A$) entonces A es consecuencia de un subconjunto finito β de α ($\beta \vDash A$).

teoría de la lógica, su repercusión se hizo sentir en muchas otras áreas centrales de la filosofía. Así, por ejemplo, Popper tras haberse resistido a toda mención de la noción de verdad, por considerarla incontrolable y metafísica, luego de los trabajos de Tarski efectuó un giro semejante al realizado por Carnap, dando a la noción semántica de verdad un lugar central en su filosofía de la ciencia. El gran impacto de la semántica de Tarski se debe en gran medida a que en ella se muestra una esclarecedora vinculación entre los signos lógicos (conectivos y cuantificadores) y la noción de verdad. En efecto, Tarski mostró cómo las mismas reglas que lógicos anteriores (como Post, Wittgenstein, Skolem y Gödel), presuponiendo la noción de verdad, habían usado para explicar el significado de los signos lógicos, podían también ser usadas para clarificar la noción de verdad (con una precomprensión de los signos lógicos).

IV. EL ENFOQUE SINTÁCTICO

Para comprender mejor el sentido de esta evolución carnapiana conviene presentar con cierto detalle su enfoque anterior de la noción sintáctica de consecuencia y las razones que lo llevaron a abandonarla como noción a tomar en cuenta para definir el área temática de la lógica.

La siguiente es posiblemente la definición más simple de la noción sintáctica (metalingüística) de consecuencia (que Carnap llamó relación de derivabilidad).

(Def. 3) Un enunciado A del lenguaje $\mathbb{L}$ es una consecuencia sintáctica del conjunto α de enunciados de $\mathbb{L}$, que abreviaremos por $\alpha \vdash A$, si y sólo si hay en $\mathbb{L}$ una secuencia finita $A_1 \ldots A_n$ de enunciados de $\mathbb{L}$, tal que $A_n = A$ y cada uno de los A_i de la secuencia es o bien un axioma de $\mathbb{L}$ o es un elemento de α o bien se sigue de enunciados que le preceden en la secuencia en función de las reglas primitivas de inferencia de $\mathbb{L}$ (de la secuencia $A_1 \ldots A_n$ se dice que es una *derivación* de la conclusión A).

En la definición anterior se entiende por axioma a todo enunciado del lenguaje $\mathbb{L}$, que por expresar lo que intuitivamente sería una verdad lógica (como lo es en la lógica clásica el enunciado (condicional) de $\mathbb{L}$ que expresa la *ley* del *modus ponens:* «$(((A \supset B) \wedge A) \supset B))$», que puede introducirse en *cualquiera* de las secuencias que constituyen una derivación. Por regla (primitiva) de inferencia se entiende a toda cláusula condicional del metalenguaje de $\mathbb{L}$ que permite introducir en una derivación el enunciado de $\mathbb{L}$ que es la conclusión de la regla si en la parte precedente de la derivación se encuentran el o los enunciados (de $\mathbb{L}$) que figuran como premisas de la regla. Así, por ejemplo, la regla del *modus ponens:* «De $(A \supset B)$ y A se sigue B», que *no* es, como la anterior ley del *modus ponens,* un enunciado (condicional) del lenguaje $\mathbb{L}$ sino un enunciado condicional de metalenguaje de $\mathbb{L}$ que permite introducir la

conclusión B en *toda* derivación en la que figuren previamente tanto el enunciado $(A \supset B)$ como el enunciado A[4].

A partir de la definición anterior puede fácilmente probarse que la relación sintáctica de consecuencia $\vdash$ cumple las siguientes tres propiedades, análogas a las indicadas en $(\models 1)$, $(\models 2)$ y $(\models 3)$ para la relación semántica de consecuencia $\models$:

$(\vdash 1)$ Reflexividad generalizada: $\alpha \vdash A$ si $A \in \alpha$.
$(\vdash 2)$ Corte: si $\alpha \vdash B$ y $\alpha \cup \{B\} \vdash A$ entonces $\alpha \vdash A$.
$(\vdash 3)$ Monotonía: si $\alpha \vdash A$ entonces $\alpha \cup \beta \vdash A$.

Conviene destacar que todas estas propiedades las cumple toda noción de consecuencia sintáctica cualquiera sea el conjunto de axiomas y el conjunto de reglas primitivas de inferencia seleccionados para identificarla (y aunque tales conjuntos sean vacíos).

Así como el caso límite en que el conjunto α de premisas es vacío da lugar a la noción de verdad lógica con referencia a la relación semántica de consecuencia, el mismo caso límite para la noción sintáctica de consecuencia, da lugar a la noción de tesis (teorema) de un sistema lógico axiomático. Un enunciado A es un teorema $\vdash A$ (abreviatura de $\emptyset \vdash A$) si y sólo si hay una secuencia como la anterior cuyos elementos son sólo axiomas o enunciados que se siguen de los precedentes en la secuencia en función de las reglas primitivas de inferencia.

Así como la noción de consecuencia semántica y la de verdad lógica dependen (en su caracterización) de las nociones de interpretación y de verdad (en particular de la satisfacción de condiciones como (i)...(v) que especifiquen las condiciones de verdad de los distintos tipos de enunciados del lenguaje), la noción sintáctica de consecuencia depende de las nociones de axioma y regla primitiva de inferencia, ya que la verdad de una afirmación sintáctica de consecuencia $\alpha \vdash A$ depende de cuáles sean los enunciados elegidos como axiomas y cuáles sean las reglas primitivas de inferencia seleccionadas al caracterizar el sistema axiomático.

En muchos casos (cuando el lenguaje incluye un signo de condicional, como '$\supset$', que satisface el llamado metateorema de la deducción:

(MD) Si $(\alpha \cup \{A\}) \vdash B$ entonces $\alpha \vdash (A \supset B)$),

y su conversa (que depende de la presencia de la regla del *modus ponens*) puede darse una definición alternativa de la lógica (que aquí es tanto como la relación sintáctica de consecuencia $\vdash$) en cuestión, identificándola con

4. Si bien no está excluida la posibilidad de que en una regla primitiva de inferencia figure un número infinito de premisas (todas las cuales tendrían que figurar en una derivación para poder introducir en ella la conclusión de la regla), la presencia de tales reglas es vacua, ya que por el carácter finito que (Def. 3) impone a cada derivación, nunca se dará la condición para usarlas. Por el contrario, debe considerarse que a los efectos de (Def. 3) están *excluidas* las reglas con condiciones negativas (del tipo «Si en la derivación figura A *pero no* B, puede introducirse C en la derivación»), características de las lógicas (no deductivas) no monótonas, cuya presencia impediría probar la monotonía de la relación de consecuencia sintáctica $(\vdash .3)$.

el conjunto de los teoremas (el conjunto de los enunciados A para los que se cumple: $\vdash$ A), ya que en las condiciones indicadas la relación sintáctica de consecuencia es definible a partir de la noción de teorema.

V. LA CUESTIÓN DE LA PRIMACÍA: SEMÁNTICA *VERSUS* SINTAXIS

Así como la definición semántica (Def. 2.0) (al igual que su equivalente (Def. 2.1)) pretende ser una reconstrucción precisa de la noción intuitiva de consecuencia (definitoria de la lógica) contenida en (Def. 1.0) (o su equivalente intuitivo (Def. 1.1)), lo mismo sucedió con la definición (Def. 3). En efecto, la definición de consecuencia sintáctica fue presentada por Carnap en sus primeros trabajos como la reconstrucción racional precisa de la noción intuitiva de consecuencia. Sin embargo, la existencia de dos nociones precisas (la sintáctica y la semántica), en principio completamente diferentes, ya que en cada una de ellas figuran esencialmente pares de conceptos muy distintos (funciones de interpretación y verdad en la definición semántica, y axiomas y reglas de inferencia primitivas en la definición sintáctica) de una misma relación intuitiva plantea varios interrogantes técnicos y filosóficos: ¿a través de cuál de las relaciones, la semántica o la sintáctica, debe llevarse adelante la tarea específica de la lógica en su empeño en reconstruir con precisión la noción intuitiva de consecuencia?, ¿debe hacerlo buscando identificar axiomas y reglas de inferencia primitivas o por el contrario debe hacerlo especificando nociones de interpretación y de verdad?

En la tarea lógica actual los dos enfoques son, en cierto modo, objeto de igual interés teórico. Si bien en la presentación original de muchas lógicas se sigue el enfoque semántico y en otras el enfoque sintáctico, hay una suerte de acuerdo tácito que considera que la tarea del lógico no se encuentra concluida hasta que para una misma lógica no se ha conseguido una presentación coincidente desde ambos enfoques. Esto significa lo siguiente. Supongamos que una específica relación de consecuencia sintáctica $\vdash$ es caracterizada a partir de la enumeración de un conjunto Ax de axiomas y de un conjunto $\mathbb{R}$ de reglas de inferencia primitivas y que una específica relación de consecuencia semántica $\models$ es caracterizada a partir de un conjunto $\mathbb{I}$ de funciones de interpretación y un conjunto $\mathbb{V}$ de cláusulas para la noción de verdad de los enunciados, entonces ambos enfoques son coincidentes (caracterizan de manera diferente una misma lógica) cuando puede probarse en el metalenguaje el siguiente enunciado de correlación:

(Corr. 1) Para todo conjunto α de enunciados de $\mathbb{L}$ y para todo enunciado A de $\mathbb{L}$: $\alpha \vdash$ A si y sólo si $\alpha \models$ A,

que para el caso particular en que se tienen en mira las nociones de teorema y verdad lógica más que las relaciones de consecuencia el enunciado anterior se convierte en:

(Corr. 2) Para todo enunciado A de $\mathbb{L}$: $\vdash$ A si y sólo si $\vDash$ A.

Ambas afirmaciones significan (si bien (Corr. 1) es más general que (Corr. 2)) que los dos enfoques producen el mismo resultado: un enunciado es una conclusión axiomática-sintáctica exactamente cuando es una consecuencia semántica, lo que implica que los enunciados sintácticamente demostrables (teoremas) son exactamente las verdades lógicas. La afirmación de correlación (Corr.) es frecuentemente identificada como un «metateorema de representación» ya que cuando el punto de partida es una presentación sintáctica este último suministra la representación semántica adecuada; y a la inversa, cuando el origen es una presentación semántica, indica una representación sintáctica equivalente.

Sin embargo, el nombre más corriente para el enunciado de correlación es el de «metateorema» de completitud (– consistencia) semántica». Este nombre enfatiza la idea de que la calidad conceptual de un sistema sintáctico-axiomático se juzga desde la perspectiva semántica, ya que tal denominación indica que lo que se ha estado buscando es la representación sintáctico-axiomática de un sistema de lógica semánticamente identificado. Giros terminológicos como el indicado, muestran cómo, consciente o inconscientemente, los lógicos contemporáneos han internalizado una posición filosófica (en cuanto a la naturaleza de la lógica) que, en cierto modo, es la posición característica de Carnap en su período semántico (el posterior a los escritos semánticos de Tarski). Podemos llamar a esta posición la de la «Primacía de la Semántica sobre la Sintaxis» (en la caracterización de la lógica). Según ella son las nociones de consecuencia semántica y de verdad lógica las que identifican a cada lógica. En este sentido identificar una lógica $\mathbb{L}_i$ es tanto como especificar la relación de consecuencia semántica correspondiente: $\vDash_i$. Si además se cuenta con una relación sintáctica correspondiente $\vdash_i$, esto es, para la que se cumple la condición de correlación (Corr. 1), entonces se está en posesión de una presentación axiomática (sintáctica) de la lógica $\mathbb{L}_i$ en cuestión.

Desde la perspectiva de la primacía de la semántica la identificación de cada lógica $\mathbb{L}_i$ es la que se logra semánticamente. La presentación axiomática es sólo una representación sintáctica de ella, obtenida prescindiendo de la significatividad del lenguaje en que la lógica es formulada, pero no la identificación de la misma. La tarea del lógico es esencialmente semántica.

Por cierto, la tesis de la primacía de la semántica no pretende restar importancia a la axiomatización de una lógica, ya que el enfoque sintáctico tiene virtudes independientes que lo justifican ampliamente. Recordemos que una afirmación de consecuencia sintáctica $\alpha \vdash$ A es una aserción (metalingüística) existencial comprometida con la existencia de una secuencia finita de enunciados del lenguaje entre cuyos puntos de partida están los enunciados de α y cuyo último enunciado es precisamente A. Como la forma más natural de representar un proceso es por medio de una secuencia (en la que se ubican, siguiendo el orden temporal, los

distintos estados momentáneos que constituyen el proceso), la caracterización sintáctica es así una fiel representación de un posible proceso psicológico de inferencia que comienza con las premisas α y que se va desarrollando según las reglas de inferencia de la lógica en cuestión hasta alcanzar como punto final la conclusión buscada. Esto es, la noción sintáctica de consecuencia conserva los rasgos estructurales de la noción psicológica de inferencia.

Muy por el contrario, las afirmaciones de consecuencia semántica $\alpha \models A$, que suponen una aserción (metalingüística) universal (y no existencial como la sintáctica)[5], sólo establecen una vinculación entre las premisas y la conclusión sin ninguna referencia a secuencia alguna que pueda reproducir algún tipo de proceso (psicológico) inferencial. Tales aserciones son vistas, desde la perspectiva de la primacía de la semántica, como el control de calidad que la lógica realiza al comparar el origen (premisas) y el punto final de la secuencia en que se constituye la derivación sintáctica, que en definitiva es la reconstrucción lingüística de un posible proceso psicológico inferencial.

Además, la inteligibilidad que se logra a través de una presentación sintáctico-axiomática de una lógica amplía y complementa a la suministrada mediante el enfoque semántico. En este sentido, el lograr una presentación axiomática significa un adelanto en la comprensión psicológica del contenido conceptual de una lógica. Lo mismo sucede habitualmente con las distintas presentaciones sintácticas de una misma lógica. Por cierto, otro tanto ocurre a la inversa. Una presentación semántica aumenta y complementa usualmente el nivel de inteligibilidad que se consigue a través de las presentaciones sintácticas.

Sin embargo, y a pesar de las virtudes señaladas del enfoque sintáctico, que son pacíficamente aceptadas, hay algunas razones muy convincentes que apoyan la tesis de la primacía de la semántica.

Un argumento en esa dirección es el siguiente. En el enfoque sintáctico se prescinde completamente del significado (en el sentido de correlación entre el lenguaje y la realidad) de las expresiones del lenguaje $\mathbb{L}$ al que está referido. Los signos y expresiones del lenguaje son, desde la perspectiva sintáctica, entidades asignificativas y sus enunciados carecen de valor de verdad (porque carecen de significado). Por este motivo la elección de los principios (axiomas y reglas de inferencia primitivas) de una lógica no tiene limitación alguna. En el enfoque sintáctico somos totalmente libres de erigir en axiomas lógicos a cualquier conjunto de enunciados del lenguaje, y lo mismo sucede con la elección de las reglas de

5. El carácter existencial de la noción sintáctica de consecuencia facilita la prueba de que un enunciado *es* consecuencia de un cierto conjunto de premisas, ya que basta con mostrar la existencia de una derivación para lograr lo buscado. A la inversa, el carácter universal de la noción semántica de consecuencia facilita la prueba de que un enunciado *no es* consecuencia de un conjunto de premisas, ya que para ello basta con mostrar la existencia de una interpretación admisible en la que las premisas son verdaderas y la conclusión no lo es.

inferencia: cualquier relación sintáctica-formal entre enunciados es susceptible de ser elegida para identificar las reglas primitivas de una lógica. Resulta así que una lógica se presenta como el resultado puramente convencional de elecciones arbitrarias adoptadas sin limitación alguna. Desde este ángulo la creación e identificación de una lógica se muestra como una empresa tan libre, arbitraria y convencional como el de la creación e identificación de un juego. Un cambio de las reglas que identifican a un juego sólo produce la identificación de otro juego tan legítimo (en tanto que juego) como el anterior.

Sin embargo, así como tenemos criterios para elegir en cada oportunidad entre los distintos juegos posibles cuál o cuáles son los más aptos para alcanzar una finalidad determinada, que puede ser muy distinta en cada contexto y circunstancia diferente (no son las mismas las finalidades que han prestigiado juegos tan distintos como la lotería, el ajedrez, el bridge, el polo, etc.). Lo mismo sucede frente a los distintos cálculos posibles que resultan de selecciones de principios diferentes. La finalidad que guía la elección entre los diversos cálculos sintáctico-axiomáticos para identificar una lógica es una finalidad muy específica y bien delimitada: se trata de que los axiomas y teoremas del sistema sean verdades lógicas y de que las reglas de inferencia (primitivas y derivadas) transmitan a la conclusión la verdad de sus premisas. Un sistema sintáctico-axiomático que no cumpla con esta condición no identifica un sistema lógico, y cualquier otra virtud que pueda tener es ajena a la lógica. Mirando las cosas de esta sensata manera, y considerando que la condición impuesta (en función de la finalidad por antonomasia de la lógica) es la que caracteriza al enfoque semántico, es como se configura uno de los caminos por los que se justifica la tesis de la primacía de la semántica.

Hay otras razones que conducen al mismo resultado. Cuando se indaga por la razón por la que ciertos enunciados expresan verdades lógicas y por la que las reglas de inferencia que preservan la verdad de su conclusión cuando sus premisas son verdaderas son identificadas como reglas *lógicas* de inferencia (que es tanto como indagar por la diferencia entre la lógica y otras disciplinas científicas), una respuesta muy antigua y sensata señala que un enunciado expresa una verdad lógica cuando su verdad puede determinarse recurriendo exclusivamente al significado de los signos lógicos que figuran en el enunciado en cuestión (o en los enunciados que figuran en la regla de inferencia, según sea el caso). Se entiende aquí que los signos lógicos, que los medievales llamaban «sincategoremáticos», son aquellos que no denotan entidad alguna de la realidad (aquellos que están fuera del ámbito de las funciones de interpretación $|\quad|_i$ antes mencionadas). Esta indicación, puramente negativa, no nos dice, sin embargo, cómo se logra la especificación del significado de los signos lógicos.

La respuesta positiva deriva de una de las tesis centrales de la filosofía de la lógica de Frege, aquélla según la cual la especificación del significado de un enunciado se obtiene mediante la explicitación de las condi-

ciones en que el enunciado es verdadero (el significado de un enunciado son sus condiciones de verdad). Desde este enfoque el significado de un signo lógico se identifica cuando se explicita cómo este último contribuye al significado de los enunciados de los que forma parte. Indicando las condiciones de verdad de los enunciados en los que interviene un signo lógico y cómo incide su presencia en las condiciones de verdad de los enunciados en los que figura se identifica (indirectamente) su significado. La definición de un signo lógico es siempre una definición de las que B. Russell llamó «definiciones en uso» (aquéllas en las que el significado de un signo se logra indicando su uso en el contexto de un enunciado). Nótese que las anteriores cláusulas (ii) hasta (v) de la caracterización de la noción de verdad para nuestro lenguaje $\mathbb{L}$, tienen como función precisamente exponer, en la forma indicada (como definición en uso), el significado de los signos lógicos de negación «$\neg$», de conjunción «$\wedge$», de disyunción «$\vee$» y de condicional (material) «$\supset$» del lenguaje modelo $\mathbb{L}$. Estas cláusulas suelen presentarse en los textos de lógica, para los conectivos proposicionales, en la forma tabular de las llamadas «tablas de verdad».

Resumiendo la argumentación anterior resulta que: a) la delimitación de las leyes (verdades) y reglas de inferencia de la lógica se hace en base al significado de los signos lógicos, b) el significado de los signos lógicos se especifica en las cláusulas que definen la noción de verdad de un lenguaje. Pero como tanto la noción de verdad como las cláusulas que la caracterizan son los rasgos esenciales del enfoque semántico, se concluye naturalmente la primacía del enfoque semántico sobre el enfoque sintáctico en la identificación del área específica de la lógica.

VI. LOS ENFOQUES SINTÁCTICOS Y SEMÁNTICOS EN LA HISTORIA

Sin perjuicio de la validez de lo anterior, es dable reconocer que el enfoque sintáctico fue la forma históricamente primera usada en la presentación de los sistemas de lógica. Ella fue la única vigente hasta bastante entrado el presente siglo. Indudablemente, sus creadores hicieron usualmente comentarios de naturaleza semántica, pero la presentación oficial siguió siempre los carriles del enfoque sintáctico.

El enfoque sintáctico fue el usado por Aristóteles en la presentación de sus sistemas lógicos (la lógica del Silogismo Categórico y la lógica del Silogismo Modal). Así, es correcto interpretar la tarea llevada a cabo por Aristóteles al presentar la lógica del silogismo categórico como la especificación de un conjunto de reglas primitivas de inferencia. Las reglas para la lógica del silogismo categórico requieren, en primer término, reglas con una única premisa, conocidas en la lógica escolástica como «Inferencias Inmediatas», y luego reglas con dos premisas llamadas «Inferencias Mediatas» que comprenden los distintos esquemas inferenciales típicos de la silogística (los silogismos indicados en los versos medievales: Barbara, Celarent, Darii, Ferio, etc.). Aristóteles mostró cómo podían

reducirse las reglas primitivas de inferencia tomando sólo como reglas mediatas los silogismos de la primera figura. Naturalmente Aristóteles admitió que una inferencia puede partir de un conjunto de premisas con más de dos enunciados, son sólo sus reglas primitivas las que no tienen nunca más de dos premisas. De este modo sus reglas de inferencia identifican una noción sintáctica de consecuencia $\vdash_s$ (donde el subíndice «s» figura para recordar que se trata de la noción de consecuencia que identifica la lógica del silogismo categórico)[6].

La presentación de la lógica del silogismo categórico por medio de una relación sintáctica de consecuencia $\vdash_s$ definida según la definición (Def. 3) a partir de un conjunto de reglas primitivas de inferencia, en pleno acuerdo con la forma usada en la escolástica, muestra algunos rasgos que importa señalar. En ella no hay axiomas (el conjunto de los axiomas con que se define $\vdash_s$ es vacío), además no hay en ella reglas que, como el metateorema de la deducción (MD), permitan disminuir el número de las premisas de una inferencia hasta alcanzar el conjunto vacío de premisas. Por estas dos razones no hay teoremas en la lógica del silogismo categórico (el conjunto de los A para los que se cumple $\vdash_s$ A es vacío). Este fenómeno es el reflejo sintáctico de la observación de von Wright de que la noción de verdad lógica fue desconocida para Aristóteles:

> It seems to me [...] that the notion of logical truth is unknown to Aristotle. This is not necessarily to blame Aristotle of ignorance. It is an interesting question, to what extent logic can be developed independently of the ideal of logical truth (von Wright, 1957, 21).

Lo anterior no significa que Aristóteles no identificó ningún enunciado como lógicamente verdadero, ya que es ampliamente conocida su discusión y defensa, en la *Metafísica*, de los principios lógicos de no contradicción y del tercero excluido. Es más, su famosa definición de la noción de verdad como correspondencia con la realidad figura precisamente en las páginas en las que discute los principios indicados. Con este alcance debiera decirse que Aristóteles llegó a su noción de verdad por haber detectado dos verdades lógicas. Creo que la importancia de la observación de von Wright deriva de la independencia teórica y conceptual entre la lógica de Aristóteles y las dos verdades lógicas por él detectadas y discutidas. Por un lado está su lógica, caracterizada sintácticamente, y por otro, sin conexión intrasistemática, dos verdades lógicas aisladas. Si Aristóteles hubiera tenido una noción general de la noción de verdad lógica habría seguramente detectado que ellas no son sólo dos y se habría

6. En la presentación del texto se sigue la presentación tradicional, no obstante la diferente opinión de Lukasiewicz, para quien cada silogismo no es una regla de inferencia con tres enunciados (dos premisas y la conclusión) sino un único enunciado condicional en el que el antecedente es la conjunción de las premisas y el consecuente es la conclusión del silogismo. Se adopta el enfoque clásico por estar convencido de la legitimidad de las críticas de C. H. von Wright al enfoque de Lukasiewicz (ver von Wright, 1957, 20).

percatado que aún con las únicas cuatro formas de enunciados tematizadas en la teoría del silogismo categórico (los enunciados de la forma A [universales afirmativos] del tipo «Todo A es B», E [universales negativos] del tipo «Ningún A es B», I [particulares afirmativos] del tipo «Algún A es B» y O [particulares negativos] del tipo «Algún A no es B») hay enunciados que expresan verdades lógicas, como es el caso de los enunciados de la forma: Todo A es A. Una consecuencia de este hecho es que si se intenta una presentación semántica de la lógica del silogismo categórico por medio de una relación de consecuencia semántica $\models_s$ seguramente se muestra que la relación sintáctica $\vdash_s$ es demasiado estrecha ya que no se podrá probar la tesis de correlación (Corr. 1), por cuanto tendremos $\models_s$ Todo A es A, pero no $\vdash_s$ Todo A es A.

También Frege usó el enfoque sintáctico en la presentación de su lógica (la hoy llamada lógica de la cuantificación de nivel superior). Frege expuso su lógica indicando un conjunto (no vacío) de axiomas y un conjunto (también no vacío) de reglas primitivas de inferencia. Si bien en la obra de Frege se encuentran las ideas semánticas más importantes con las que en el futuro se construiría la noción semántica de consecuencia correspondiente a la noción sintáctica usada por Frege, el instrumental técnico que supone la identificación semántica no existía aún en los tiempos de Frege. Recién K. Gödel en 1930 probó el teorema de correlación entre la relación sintáctica de consecuencia $\vdash_Q$ para la porción de la lógica de Frege conocida actualmente como lógica de la cuantificación (cálculo funcional de primer orden) y la relación semántica de consecuencia $\models_Q$ que fue caracterizada por Gödel con elementos que derivan básicamente de Post y Wittgenstein (como coinventores independientes de las tablas de verdad) y de T. Skolem en obras suyas de los años 1919 y 1920.

Posiblemente las primeras lógicas presentadas originariamente desde la perspectiva semántica sean las lógicas polivalentes, iniciadas a principio de la década del treinta, ya que en ellas el significado de los signos lógicos es identificado por un método de matrices tabulares análogas a las tablas de verdad y su noción de consecuencia es caracterizada por una definición generalizada del tipo de (Def. 2.0). Por esta razón una de las tareas más absorbentes, con relación a las lógicas polivalentes, fue la de encontrar presentaciones sintáctico-axiomáticas con las que se satisfagan las condiciones de correlación (Corr. 1) y Corr. 2).

Tanto la lógica intuicionista de Heyting (1930) como las lógicas modales de C.I. Lewis, con las que se inicia el período moderno de la lógica modal, fueron concebidas usando el enfoque axiomático-sintáctico. Aun cuando las pioneras y fundamentales indagaciones en lógica modal de G.H. von Wright fueron realizadas desde una perspectiva semántica, lo cierto es que la semántica estándar de la lógica modal debió esperar hasta los trabajos de S. Kripke, S. Kanger y J. Hintikka que en la década del sesenta consolidan el aparato teórico para la presentación semántica de estas lógicas. Las lógicas modales de Lewis y los modelos semántico

modales de Kripke tienen una gran incidencia en el tema que nos ocupa: la caracterización del área temática de la lógica. Por este motivo dedicaremos a ellas los próximos párrafos.

VII. REFINAMIENTO DEL ENFOQUE SEMÁNTICO

Lewis consideraba que la tarea fundamental de la lógica era la de reconstruir con precisión la noción intuitiva de consecuencia lógica (Def. 1.0) y entendió que esto debía hacerse incorporando al lenguaje objeto un signo que representara la noción de consecuencia (cosa que no sucede en el lenguaje de Frege ni en el de ninguno de los que continuaron con su estilo). Es así como Lewis enriquece el lenguaje objeto, que ahora llamaremos $\mathbb{Lm}$ (lenguaje modal), con el signo para lo que llamó «implicación estricta» (para el que aquí usaremos «$\Rightarrow$») de manera tal que una expresión del tipo «($A \Rightarrow B$)», leído «A implica estrictamente a B», sea verdad cuando B es una consecuencia lógica de A[7].

Al llevar al lenguaje objeto, mediante la implicación estricta, la representación de la noción de consecuencia, que en este escrito (siguiendo el hábito actual) fue hasta ahora una noción metalingüística (tanto en el caso de $\vdash$ como en el de $\models$), suceden varias transformaciones que sutilmente introducen rasgos importantes en el análisis de la noción de consecuencia lógica.

En primer lugar, la noción de consecuencia sufre una limitación. En efecto, tanto en el caso de la relación sintáctica $\vdash$ como en el de la relación semántica $\models$ de consecuencia lo que figura a la izquierda es la referencia a (el nombre de) un conjunto α de premisas. Ese conjunto puede ser tan grande como se quiera: puede ser un conjunto infinito de enunciados. Muy por el contrario en una implicación estricta lo que figura a la izquierda de $\Rightarrow$ es un único enunciado, de modo que lo que las implicaciones estrictas pueden reconstruir no es la idea de cuándo un enunciado es consecuencia lógica de un conjunto de enunciados (premisas), sino sólo cuándo un enunciado es consecuencia lógica de otro (o mejor, cuándo la proposición expresada por un enunciado es consecuencia lógica de la proposición expresada por otro enunciado). Naturalmente cuando el conjunto α de premisas es finito entonces habrá un enunciado del lenguaje que pueda ocupar su lugar, sin pérdida de alcance conceptual, como antecedente en una implicación estricta, pero cuando α no es de ese tipo la implicación estricta es inhábil para dar cuenta de la noción de consecuencia. Por esta razón el recurso de Lewis supone una restricción finitista de la noción de consecuencia de la que la noción de consecuencia

7. En rigor de verdad $\Rightarrow$ no fue el signo primitivo con que Lewis enriqueció el lenguaje $\mathbb{L}$ para alcanzar el lenguaje modal $\mathbb{Lm}$. Su lenguaje modal estaba caracterizado por la introducción de un rombo para representar la noción modal de posibilidad, y definir luego la implicación estricta ($A \Rightarrow B$) (al estilo de (Def. 1.1)) cuando se da el caso que no es posible A en conjunción con la negación de B.

semántica $\vDash$ carece. La noción de consecuencia sintáctica es también finitista, si como lo hemos hecho en (Def. 3), siguiendo la tradición más estable, una afirmación del tipo $\alpha \vdash$ A requiere la existencia de una secuencia *finita* de enunciados (lo anterior significa que la noción de consecuencia sintáctica introducida en (Def. 3) *es compacta*). Sin embargo, la finitud de la noción de consecuencia sintáctica $\vdash$ es de muy distinta naturaleza que la finitud que la implicación estricta $\Rightarrow$ conlleva, ya que ella no impide indagar por las consecuencias de un conjunto infinito (ni para la significatividad de $\alpha \vdash$ A ni para su verdad se requiere que α sea un conjunto finito).

En segundo lugar, dado que una implicación estricta (A $\Rightarrow$ B) es un enunciado del lenguaje a la par de los que figuran como su antecedente A y su consecuente B, resulta que en el enfoque de Lewis son expresables implicaciones estrictas anidadas, esto es, implicaciones estrictas del tipo ((A $\Rightarrow$ B) $\Rightarrow$ C) o del tipo (C $\Rightarrow$ (A $\Rightarrow$ B)) en donde una implicación estricta figura en el ámbito (como parte del antecedente o del consecuente) de otra implicación estricta. Este rasgo de la implicación estricta es el reflejo de uno de los problemas filosóficos e intuitivos más oscuros de la lógica modal: la cuestión de las modalidades reiteradas. En las nociones metalingüísticas de consecuencia (sintáctica y semántica) este rasgo perturbador de la implicación estricta no existe. En los enunciados del tipo «$\alpha \vdash$ A», o del tipo «$\alpha \vDash$ A», «α» es el nombre de un conjunto de enunciados del lenguaje objeto y «A» el nombre de un enunciado del lenguaje objeto, pero como tanto «$\alpha \vdash$ A» como «$\alpha \vDash$ A» no son enunciados del lenguaje objeto sino de su metalenguaje ni tampoco son nombres de nada, su función no es nombrar cosa alguna, sino afirmar la relación de consecuencia entre el conjunto de enunciados nombrado por «α» y el enunciado nombrado por «A». Por esta razón carecen de sentido las expresiones en donde a la izquierda o a la derecha de alguno de los signos de consecuencia «$\vdash$» o «$\vDash$» aparezca una expresión que contenga alguno de tales signos. De este modo las oscuras cuestiones a que ha dado lugar el anidamiento de los condicionales estrictos carecen de sentido y no pueden plantearse para las nociones metalingüísticas de consecuencia.

Hasta ahora hemos señalado dos rasgos en que la implicación estricta está en desventaja en la comparación con las nociones metalingüísticas comentadas como reconstrucción de la noción intuitiva de consecuencia. Sin embargo, hay un aspecto en el que el balance puede serle favorable: en el enfoque de Lewis se tematiza directamente la noción de necesidad involucrada en el concepto intuitivo de consecuencia lógica. Las consecuencias y la relevancia filosófica de esta característica se ponen más claramente de manifiesto dirigiendo nuestra atención a los instrumentos semánticos que Kripke usó para dar cuenta semántica de las lógicas modales.

El instrumental usado hasta ahora para dar explicación semántica de la noción intuitiva de consecuencia estaba integrado por elementos de dos tipos: un conjunto de funciones de interpretación $| \quad |_i$ que fijan

la referencia (significado extensional) a los nombres (términos singulares) y a los predicados del lenguaje objeto, y un conjunto de cláusulas en las que se indican las condiciones para la verdad de los distintos tipos de enunciados. La forma de estas últimas cláusulas: «un enunciado X es verdadero en $|\quad|_i$ si y sólo si ...» (en las que la verdad de un enunciado es relativa a la interpretación que se esté considerando), explican la razón por la que un enunciado puede cambiar su valor de verdad (pasar de ser verdadero a ser falso o a la inversa) al cambiar la interpretación a través de la cual se juzga el valor de verdad del enunciado. También señalamos que con la noción semántica de verdad Tarski pretendió reconstruir para un lenguaje artificial (como $\mathbb{L}$) la noción aristotélica de verdad como correspondencia con la realidad (para obtener este resultado son esenciales las cláusulas del tipo de (i), esto es, las cláusulas para los enunciados atómicos, ya que por medio de ellos se inicia la «correspondencia con la realidad»). Esta última aspiración de correspondencia significa que si la realidad considerada se transforma (cambia en alguno de sus aspectos), entonces el valor de verdad de un enunciado que a ella se refiere puede cambiar reflejando los cambios ocurridos en el mundo. En otras palabras, de una noción de verdad por correspondencia referida a entidades lingüísticas (enunciados) esperamos que el valor de verdad de un enunciado pueda cambiar por dos razones distintas e independientes: a) por un cambio en la interpretación con que son entendidas las expresiones que figuran en el enunciado, o b) por un cambio en la realidad considerada. En las cláusulas al estilo de Tarski (las consideradas hasta ahora) la relatividad de la verdad con relación a la interpretación del lenguaje es explícita pero, muy por el contrario, en ellas no se prevé la relatividad a la realidad considerada, cuyos cambios también pueden incidir en el valor de verdad[8].

Al elaborar el instrumental para dar cuenta de la semántica de las lógicas modales Kripke colmó el vacío indicado, dejado por Tarski, en las cláusulas de su definición semántica de verdad. En efecto, las cláusulas usadas por Kripke responden a la siguiente forma estructural: «Un enunciado X (del lenguaje $\mathbb{L}$) es verdadero en la interpretación $|\ ,\ |_i$ frente a la realidad (mundo) M_j si y sólo si ...».

La realidad extralingüística a la que está referida un lenguaje $\mathbb{L}$, que en Kripke se llama «Estructura de Modelo», es una estructura $\mathbb{G}$ integrada por dos elementos: 1) un conjunto $\mathbb{M}$ de mundos posibles (en un ejemplo intuitivo dinámico-temporal como el anterior referido a cambios temporales de la realidad, los elementos de $\mathbb{M}$ pueden ser interpretados como los distintos estados totales de la realidad en los diferentes momentos temporales; naturalmente esa interpretación intuitiva es inadecuada cuando se pretende representar la posibilidad lógica) y 2) un ele-

8. Quizás una posible conjetura histórica para explicar esta omisión de Tarski sea que en los ejemplos que elaboró con detalle sólo figuran teorías matemáticas y lógicas, y es claramente insensato pensar en el cambio de una realidad del tipo de la que tratan esas disciplinas.

mento destacado $\mathbb{G}$ de $\mathbb{M}$ destinado a representar el mundo (la realidad) actual frente a los mundos meramente posibles (los restantes elementos de $\mathbb{M}$)[9]. La referencia al lenguaje $\mathbb{L}$ aparece en lo que Kripke llama un modelo $\mathfrak{M}_{\mathbb{G}}$ (para una estructura de modelo $\mathbb{G}$) que se consigue acoplando a la estructura de modelo $\mathbb{G}$ una función de interpretación $|\mathbb{S},\mathbb{M}_j|_i$ de dos argumentos que adjudica: 1) a cada nombre $\mathbb{S}$ (cuando $\mathbb{S}$ es un nombre de $\mathbb{L}$) frente a cada mundo $\mathbb{M}_j$ de $\mathbb{M}$ un único objeto (de $\mathbb{M}_j$), y 2) a cada predicado $\mathbb{S}$ (cuando $\mathbb{S}$ es un predicado monádico de $\mathbb{L}$) un único conjunto de objetos (de $\mathbb{M}_j$). Con estos elementos las cláusulas que definen la noción de verdad para todo modelo $\mathfrak{M}_{\mathbb{G}}$ son ahora (en lugar de las anteriores):

(i') Un enunciado atómico Pa es verdadero en la interpretación $| \ , |_i$ frente al mundo $\mathbb{M}_j$ si y sólo si $|a,\mathbb{M}_j|_i \in |P,\mathbb{M}_j|_i$ (el objeto asignado al nombre «a» por la interpretación $| \ , |_i$ para el mundo $\mathbb{M}_j$ es uno de los elementos de la clase asignada al predicado «P» por esa misma interpretación y para ese mismo mundo).

(ii') $\neg$ A es verdad en $| \ , |_i$ frente a $\mathbb{M}_j$ si y sólo si A no es verdad en $| \ , |_i$ frente a $\mathbb{M}_j$.

(iii') (A $\wedge$ B) es verdad en $| \ , |_i$ frente a $\mathbb{M}_j$ si y sólo si tanto A como B son verdad en $| \ , |_i$ frente a $\mathbb{M}_j$.

(iv') (A $\vee$ B) es verdad en $| \ , |_i$ frente a $\mathbb{M}_j$ si y sólo si A, B o ambas son verdad en $| \ , |_i$ frente a $\mathbb{M}_j$.

(v') (A $\supset$ B) es verdad en $| \ , |_i$ frente a $\mathbb{M}_j$ si y sólo si A no es verdad en $| \ , |_i$ frente a $\mathbb{M}_j$ o B es verdad en $| \ , |_i$ frente a $\mathbb{M}_j$.

(vi) (A $\Rightarrow$ B) es verdad en $| \ , |_i$ frente a $\mathbb{M}_j$ si y sólo si B es verdad en $| \ , |_i$ frente a $\mathbb{M}_k$ para todo $\mathbb{M}_k$ de $\mathbb{M}$ en el que A es verdad en $| \ , |_i$ frente a $\mathbb{M}_k$.

(vii) $\Box$A es verdad en $| \ , |_i$ frente a $\mathbb{M}_j$ si y sólo si A es verdad en $| \ , |_i$ frente a $\mathbb{M}_k$ para todo $\mathbb{M}_k$ de $\mathbb{M}$.

En la cláusula (vii) hemos usado «$\Box$A» para el enunciado que representa la afirmación de la necesidad lógica de A. Esa cláusula reproduce la idea leibniziana según la cual son lógicamente necesarios aquellos enunciados que son verdaderos en todo mundo posible. Las condiciones que (vi) fija para la verdad de una implicación estricta (A $\Rightarrow$ B) (que en la interpretación de Lewis afirma que B es consecuencia lógica de A) son las mismas que por (vii) y (v') resultan para $\Box$(A $\supset$ B) (la necesidad lógica de la implicación material).

9. Una estructura de modelo como la indicada en el texto con sólo dos elementos es instrumental suficiente para dar cuenta de la semántica del sistema modal S5 (que fue el objetivo de Kripke cuando escribió su primer artículo sobre el tema). Para dar cuenta de otros sistemas modales Kripke creó las «estructuras de modelos relacionales» en las que aparece como tercer elemento una relación $\mathbb{R}$ (relación de accesibilidad) que vincula elementos de $\mathbb{M}$. Sin embargo, la estructura más simple del texto es suficiente para nuestros fines.

En las cláusulas anteriores la noción de verdad es relativa (como se espera en una teoría de verdad por correspondencia) a la interpretación considerada de las expresiones lingüísticas y a la realidad (mundo) a que se refiere el enunciado. Para explicar el alcance de la noción de verdad cuando la usamos (como es frecuente en la vida diaria) sin relativización alguna, Kripke introduce lo que llama la *verdad de* A *en un modelo* $\mathfrak{ME}$, que se da cuando A es verdadero en el mundo destacado (actual) $\mathbb{G}$ del modelo $\mathfrak{ME}$. Se recoge así la idea intuitiva que refiere al mundo actual a los enunciados para los que no explicitamos una referencia distinta.

Con estos elementos podemos dar los siguientes refinamientos de las definiciones (Def. 2.0) y (Def. 2.1) de la relación de consecuencia semántica $\vDash$.

(Def. 2.2) Un enunciado A de $\mathbb{L}$ es consecuencia (semántica) del conjunto de enunciados α de $\mathbb{L}$ (premisas), $\alpha \vDash$ A, si y sólo si no hay ningún modelo $\mathfrak{ME}$ de $\mathbb{L}$ en el que todos los enunciados de α son verdaderos y en el que A no lo es.

(Def. 2.3) Un enunciado A de $\mathbb{L}$ es consecuencia (semántica) del conjunto de enunciados α de $\mathbb{L}$ (premisas), $\alpha \vDash$ A, si y sólo si A es verdadera en todo modelo $\mathfrak{ME}$ de $\mathbb{L}$ en el que son verdaderos todos los enunciados de α.

El refinamiento introducido en la noción de consecuencia semántica $\vDash$ en las últimas definiciones deriva básicamente de que en ellas se ha tomado explícitamente en cuenta, además de la relativización de la noción de verdad a la interpretación adoptada de las expresiones lingüísticas, la relativización con relación a la realidad considerada, dando así expresión más acabada al requerimiento de correspondencia incluido en la noción intuitiva de verdad. También ha cambiado la explicación de las nociones modales (de necesidad e imposibilidad), asociadas a las nociones intuitivas de consecuencia y verdad lógica. En (Def. 2.0) y (Def. 2.1) la explicación reposaba en una cuantificación universal sobre las funciones de interpretación admisibles. Por el contrario, en las nuevas definiciones, tales nociones, dependen de un enfoque más leibniziano que requiere una cuantificación sobre mundos posibles.

Finalmente, lo anterior significa una transformación sustancial que afecta a las nociones mismas de consecuencia semántica y verdad lógica. Esto es así por cuanto, en las últimas definiciones, para que un enunciado A sea consecuencia semántica de un conjunto α de premisas, $\alpha \vDash$ A (al igual que para que A sea una verdad lógica: $\vDash$ A), se requiere una doble cuantificación metalingüística: una sobre todas las interpretaciones admisibles y otra sobre todos los mundos posibles (esta doble cuantificación está involucrada en la explícita cuantificación contenida en las últimas definiciones sobre todos los modelos). Un efecto, señalado con poca frecuencia, de esta doble cuantificación es que ella evita la posibilidad (que en principio dejan abiertas las definiciones (Def. 2.0) y (Def. 2.1))

de la existencia de verdades lógicas no lógicamente necesarias y de verdades lógicamente necesarias que no son identificadas como verdades lógicas. La necesidad de tomar en cuenta esta doble cuantificación en la explicación de las nociones centrales de la lógica, con frecuencia no es advertida, porque en muchas oportunidades hay una correspondencia biunívoca entre lo que es un mundo posible y lo que es una interpretación admisible. Así por ejemplo, las cuatro alternativas de una tabla de verdad estándar para dos enunciados:

A	B
V	V
F	V
V	F
F	F

admite una doble lectura: 1) como las cuatro interpretaciones distintas que dos enunciados pueden recibir, o 2) como los cuatro tipos de mundos posibles que pueden fijar el valor de verdad de un par de enunciados. Si bien el tema de la correlación entre mundos posibles y valuaciones admisibles es de gran importancia técnica y filosófica, no diremos nada más sobre él, por cuanto es más propio de un ensayo sobre lógica modal que de uno sobre la lógica en general.

VIII. UNA DIFICULTAD A LA PRIMACÍA DE LA SEMÁNTICA

Con las nuevas interpretaciones (Def. 2.2) y (Def. 2.3) de la relación semántica de consecuencia $\models$ podemos retornar al análisis de la tesis de la prioridad de la semántica. La circunstancia de que el enfoque sintáctico sea el que más se aproxima a la noción psicológica intuitiva de inferencia, y de que ese enfoque reconstruya con gran precisión lo que se ha hecho durante siglos de trabajo en el área de la lógica no invalida la pretensión de la tesis de la primacía de la semántica. En efecto, lo que tal tesis pretende es suministrar una explicación precisa de los criterios utilizados para evaluar el trabajo de los lógicos: es en la preservación de la noción semántica de verdad en el paso de las premisas de una regla sintáctica de inferencia a su conclusión como se aprecia la calidad de la regla. Por esta razón y no por consideraciones históricas es que se justifica erigir la noción semántica de consecuencia en el criterio definitorio de la lógica. Sin embargo, la siguiente es una argumentación que invalida la tesis de la primacía de la semántica en la definición general de la lógica.

Hay enunciados respecto de los cuales es lugar corriente reconocer que carecen completamente de valores de verdad. Paradigma de ellos son los enunciados usados por el legislador para prescribir normativamente

la conducta de los súbditos de un país, como, por ejemplo, los enunciados que usa un legislador para exigir a los súbditos el pago de un impuesto o para autorizar la importación de cierto tipo de mercancías. Sus enunciados no están destinados a describir algo que ocurre en el mundo, su función es prescribir una forma de comportamiento como debida (o indebida) o autorizar (permitir) ciertos modos de actuar. De los dichos del legislador puede afirmarse que son justos o injustos, convenientes o inconvenientes, etc., pero carece de sentido predicarles verdad o falsedad. Sólo de los enunciados descriptivos tiene sentido la predicación de valores de verdad, ya que afirmar que un enunciado es verdadero significa que en el mundo acaece lo que el enunciado describe, y afirmar que es falso supone que en el mundo no sucede lo que el enunciado describe; de modo que en ambos casos (para tener un valor de verdad) el enunciado tiene que ser descriptivo, tiene que describir el acontecer de un hecho (que puede o no darse). Nada de esto sucede con las normas del derecho positivo (el derecho creado por los hombres) pues ellas prescriben comportamientos sin describir el acontecer de ningún hecho.

El hecho de la existencia de enunciados significativos carentes de valores de verdad plantea una de las dificultades filosóficas más serias de la lógica deóntica (interpretada como lógica de normas). Esta dificultad se la conoce como el dilema de Jörgensen cuyo alcance conceptual es el siguiente: si las nociones de la lógica sólo pueden definirse en función de valores de verdad entonces no es posible una lógica de normas, y a la inversa: si tal lógica es posible las nociones de la lógica no dependen de los valores de verdad de los enunciados, luego o no es posible una lógica de normas, o las nociones de la lógica son independientes de los valores semánticos de verdad y falsedad.

La importancia de esta dificultad deriva de las siguientes dos observaciones:

1. Del hecho que los signos lógicos (conjunción, disyunción, negación, etc.) se usan significativamente en los enunciados que expresan normas sin diferencia detectable con relación al modo en que se los usa en los enunciados descriptivos, se sigue que el significado semántico (el que deriva de las condiciones de verdad de los enunciados en los que los signos lógicos figuran) no es el único relevante para la significatividad de los signos lógicos y hace posible la hipótesis de que sea otro tipo de significatividad la que unifica el significado de los signos lógicos en los enunciados descriptivos y en los prescriptivos.

2. La forma en que a diario entendemos las expresiones normativas indica que confiamos que entre ellas hay relaciones lógicas. Nuestra intelección del lenguaje normativo indica que algunas normas son consecuencia deductiva de otras. En efecto, si de un enunciado normativo cuantificado universalmente (como lo son las leyes generales promulgadas por el legislador) no se pudiera concluir deductivamente la totalidad de los enunciados normativos que refieren el contenido normativo de la ley gene-

ral a cada uno de los súbditos (es decir si la regla de ejemplificación universal no fuera aplicable a las normas), entonces ningún súbdito tendría su conducta regulada por las leyes generales y éstas se convertirían en meros juguetes vistosos sin la significatividad normativa con que incuestionablemente son entendidas en la vida social.

La dificultad indicada de la lógica deóntica expuesta en el dilema de Jörgensen cuyo primer cuerno (no es posible una lógica deóntica de normas) se muestra como incuestionablemente falso por las razones expuestas es un desafío sólo para la tesis de la primacía de la semántica. A la inversa constituye una virtud de los otros enfoques (el sintáctico y el abstracto, que presentaremos a continuación) la circunstancia de que en ellos los enunciados descriptivos y los prescriptivos admiten sin dificultad un tratamiento en paridad de condiciones.

IX. EL ENFOQUE GENERAL ABSTRACTO

El hecho de que tanto las nociones sintácticas como las nociones semánticas de consecuencia sean importantes candidatos para la reconstrucción de la noción intuitiva y de que ambas son en definitiva nociones de consecuencia deductiva, a pesar de las enormes diferencias que las separan, hace pensar en la existencia de rasgos comunes que ambos tipos de nociones comparten, y que, quizás, ellos suministren la pista para la caracterización de una noción general de consecuencia de la cual tanto el enfoque sintáctico como el semántico no sean más que especificaciones diferentes. A responder a este interrogante y lograr así una noción general abstracta de consecuencia están destinados algunos trabajos de Tarski de la década del treinta, los primeros de los cuales (Tarski, 1930a y 1930b) preceden al artículo en donde Tarski presenta su definición semántica (Tarski, 1936). Un rasgo fundamental del enfoque abstracto es que en él no se intenta, como en los otros enfoques (el sintáctico y el semántico), caracterizar la noción de consecuencia por medio de esquemas de definición (como los expuestos anteriormente), sino señalando las propiedades generales que identifican a toda noción de consecuencia (deductiva). Esto significa que la noción de consecuencia se toma como un término primitivo (no definido) sujeto a varios axiomas que identifican sus propiedades esenciales.

Para axiomatizar la noción de consecuencia Tarski usó una *función de consecuencia* «Cn» que pertenece al metalenguaje de un lenguaje $\mathbb{L}$, esto es, una función que aplicada a un conjunto de enunciados α de un lenguaje $\mathbb{L}$ identifica otro conjunto de enunciados $Cn(\alpha)$ de $\mathbb{L}$ como el conjunto de la totalidad de las consecuencias de α. El que no se represente la noción de consecuencia como una *relación* (como se hizo en el enfoque sintáctico y en el semántico) es sólo una diferencia técnica sin importancia conceptual, ya que el siguiente esquema permite pasar tanto

de una función de consecuencia a la correspondiente relación de consecuencia como a la inversa:

AεCn(α) si y sólo si $\alpha \vdash$ A (o $\alpha \vDash$ A en su caso).

Los siguientes cuatro axiomas exponen las propiedades esenciales (comunes) a toda noción de consecuencia desde la perspectiva de los primeros trabajos de Tarski:

(Cn.1) $\alpha \subseteq$ Cn(α) (Inclusión)
(Cn.2) Cn(α) = Cn(Cn(α)) (Idempotencia)
(Cn.3) Si $\alpha \subseteq \beta$ entonces Cn(α) $\subseteq$ Cn(β) (Monotonía)
(Cn.4) Si AεCn(α) entonces hay un $\beta \subseteq \alpha$ finito tal que AεCn(β) (Compacidad)

Los dos primeros axiomas exponen propiedades muy intuitivas: que todo enunciado de un conjunto está entre las consecuencias de ese conjunto (Inclusión) y que las consecuencias de las consecuencias de un conjunto de enunciados son consecuencias del conjunto de partida (Idempotencia). El tercero de los axiomas (Monotonía) indica que si algún enunciado es consecuencia de un conjunto de premisas α, él seguirá siendo consecuencia de cualquier ampliación β del conjunto de premisas; en otras palabras, que al agregar enunciados a un conjunto de premisas no se pierde ninguna de sus consecuencias. Este axioma responde a la idea de que las premisas de una inferencia deductiva son condición de garantía suficiente de sus consecuencias, de modo que cuando ellas están presentes, aunque estén acompañadas por otros enunciados, sus consecuencias no se pierden por la presencia de tales premisas adicionales. Esta es una propiedad esencial de toda noción de consecuencia *deductiva* (que es la que estamos tematizando en este ensayo), es decir, que la monotonía es esencial para la idea de deducción más que para la de consecuencia en general. Si estuviéramos indagando una noción de consecuencia probable seguramente no esperaríamos que ella sea monótona. Uno de los focos de investigación actual en el área de la llamada Inteligencia Artificial está centrada en la indagación de nociones de consecuencia no monótonas, y por lo tanto, no deductivas. Pero como nuestro interés en este ensayo está circunscripto a la lógica deductiva, el postulado de monotonía tendrá que ser satisfecho por toda noción de consecuencia cuya caracterización perseguimos.

El cuarto axioma (Compacidad) impone a la noción de consecuencia una restricción finitista que, como señalamos anteriormente, tiene su origen en una sólida tradición vinculada a la noción sintáctica de consecuencia, pero que es ajena al enfoque semántico. En este sentido él no expresa una propiedad que pueda atribuirse a toda noción de consecuencia deductiva por más deseable y atractiva que resulte su exigencia. Quizás por esta razón, y a pesar de su postulación por Tarski, ella no figura entre las condiciones actualmente exigidas en la caracterización general

abstracta de la noción de consecuencia deductiva. Por ello los axiomas con que se caracteriza tal noción son sólo los tres primeros.

Una de las investigaciones más importantes y representativas de la lógica de este siglo, la memoria de G. Gentzen (Gentzen 1934) sobre la deducción lógica, está estructurada desde la perspectiva del enfoque abstracto. Es más, el reconocimiento actual de la importancia de este enfoque deriva en mucha mayor medida de la repercusión de la obra de Gentzen que de la de Tarski sobre el tema.

Gentzen, como Tarski, adoptó como primitiva la noción de consecuencia deductiva y la caracterizó por el método axiomático (axiomas y reglas primitivas de inferencia). No obstante, su formulación difiere de la de Tarski en algunos aspectos que conviene señalar, no sin advertir que el contenido conceptual de la obra de ambos autores es sustancialmente el mismo.

Gentzen caracterizó una *relación* de consecuencia lógica, y no como Tarski una *función* de consecuencia. Ésta, como ya indicamos, es una diferencia totalmente accidental (para representarla usó la flecha «→»). La segunda diferencia es sustancial; Gentzen generaliza la estructura de las relaciones de consecuencia sintáctica $\vdash$ y semántica $\models$ permitiendo la presencia de conjuntos de enunciados tanto a la izquierda como a la derecha de la flecha, de modo que los enunciados básicos (que llama secuentes) son del tipo «$\alpha \rightarrow \beta$» (llamando «prosecuente» al que está a la izquierda de la flecha y «postsecuente» al de la derecha)[10]. En realidad esta diferencia hace que lo que la flecha representa no sea la relación de consecuencia sino una relación más general que Carnap (1943) llamó «lógical involution»[11]. Intuitivamente se espera tener $\alpha \rightarrow \beta$ si por lo menos uno de los elementos de β es verdadero cuando todos los elementos de α lo son. La relación de consecuencia lógica esta representada por el caso especial de los secuentes en que el postsecuente es un conjunto unitario (donde figura un y sólo un enunciado).

Otra diferencia es que la flecha «→» no es un signo del metalenguaje (del lenguaje donde figuran los signos lógicos) como lo es el signo de función de consecuencia «Cn» de Tarski, sino un signo incorporado al mismo lenguaje objeto en el que se encuentran los signos lógicos. Sin embargo, los únicos enunciados que son axiomatizados en su cálculo de secuentes son los secuentes y sólo indirectamente los enunciados corrientes de una lógica estándar. Es decir, los únicos enunciados axiomatizados son los

10. En el texto los elementos α y β que integran un secuente $\alpha \rightarrow \beta$ son conjuntos de enunciados. En la obra de Gentzen son secuencias (conjuntos ordenados) de enunciados. Esta diferencia, que probablemente sea un resabio de las secuencias que intervienen en la caracterización de la noción sintáctica de consecuencia, la hemos cancelado por cuanto en su presentación Gentzen introduce las postulaciones que hacen irrelevante el orden de los elementos de las secuencias, las que en definitiva se comportan como simples conjuntos.

11. Aparentemente Carnap descubrió, independientemente de Gentzen, las propiedades y la importancia de la generalización que representa la relación de «logical involution» con respecto a la de consecuencia lógica.

enunciados del tipo «$\alpha \rightarrow \beta$» donde los elementos tanto de α como de β son enunciados de un lenguaje lógico estándar (en el caso de Gentzen son los enunciados del lenguaje de cuantificación). Nunca la flecha «$\rightarrow$» figura en el prosecuente α ni en el postsecuente β.

En este aspecto Gentzen sigue el camino de Lewis, ya que sus secuentes son, como las implicaciones estrictas de Lewis, enunciados del lenguaje objeto (destinados en ambos casos a dar cuenta de la relación de consecuencia deductiva). Sin embargo, hay por lo menos tres muy importantes diferencias con el encuadre de Lewis: 1) Los secuentes no forman parte, como las implicaciones estrictas, de un lenguaje modal, 2) Los secuentes permiten dar cuenta de la situación en que hay una pluralidad de premisas ya que lo que figura a la izquierda de la flecha es un conjunto de enunciados y no un único enunciado como requiere la implicación estricta de Lewis, y 3) la diferencia más sustancial con Lewis es que no se plantea el problema de las implicaciones estrictas anidadas ya que lo que tiene que figurar tanto como prosecuente como postsecuente son conjuntos de enunciados de un lenguaje de lógica estándar pero los secuentes mismos no son enunciados de ese tipo.

No vamos a entrar en los detalles de la exposición de Gentzen; no obstante conviene señalar que su axiomatización de los secuentes es sustancialmente[12] coincidente con la de Tarski para su función de consecuencia. Los axiomas de Gentzen para la relación de consecuencia (no para la relación de «logical involution») son los tres: (Reflexividad Generalizada), (Corte) y (Monotonía) que se indicaron anteriormente para las relaciones sintácticas $\vdash$ y semánticas $\models$ de consecuencia. Esto es, ellos son:

($\rightarrow$.1) $\alpha \rightarrow \{A\}$ si $A \varepsilon \alpha$ (Reflexividad Generalizada)
($\rightarrow$.2) Si $\alpha \rightarrow \{B\}$ y $\alpha \cup \{B\} \rightarrow \{A\}$
entonces $\alpha \rightarrow \{A\}$ (Corte)
($\rightarrow$.3) Si $\alpha \rightarrow \{A\}$ entonces $\alpha \cup \beta \rightarrow \{A\}$ (Monotonía)

De este modo tanto en el caso de Tarski como en el de Gentzen la lógica (cada lógica) está caracterizada por cada función (relación) abstracta de conjuntos de enunciados de un lenguaje $\mathbb{L}$ a conjuntos de enunciados del mismo lenguaje que satisfaga las condiciones incluidas en las axiomatizaciones indicadas.

Para conseguir cada lógica en particular sólo hay que agregar a los axiomas generales de la noción de consecuencia: (Inclusión), (Idempotencia) y (Monotonía), otros que indiquen el comportamiento de los signos lógicos en el contexto de una función (relación) de consecuencia.

12. En el texto se habla de una «sustancial» equivalencia entre los tres primeros axiomas de Tarski y las tres postulaciones de Gentzen. Sin embargo, para interderivar mutuamente los postulados de Tarski con los de Gentzen es necesario, como lo ha observado David Makinson, generalizar la regla de corte ($\rightarrow$.2) de la siguiente manera: si $\alpha \rightarrow \{B\}$ para todos los $B \varepsilon \beta$ y $\alpha \cup \beta \rightarrow \{A\}$ entonces $\alpha \rightarrow \{A\}$ (en donde lo que se «corta» no es un enunciado B sino un conjunto β de enunciados).

Así, para conseguir, por ejemplo, la lógica proposicional clásica (para los signos lógicos usados en los ejemplos anteriores), basta con agregar (en una presentación que no sigue la forma de la lógica de secuentes de Gentzen, sino que representa más bien una versión que reproduce los esquemas de «deducción natural» creados también por el propio Gentzen en el trabajo comentado) los siguientes axiomas (a continuación un conjunto $\{A_1,\ldots,A_n\}$ se escribirá simplemente $A_1,\ldots,A_n$; se usará, además la constante de falsedad (lógica) « ⊥ » con el axioma que se indica):

(A.⊥) $\bot \rightarrow A$

De Introducción	De Eliminación
(I.∧) $A,B \rightarrow (A \wedge B)$	(E.∧.1) $(A \wedge B) \rightarrow A$
	(E.∧.2) $(A \wedge B) \rightarrow B$
(I.v.1) $A \rightarrow (A \vee B)$	(E.v) $\alpha \cup (A \vee B) \rightarrow C$ si $\alpha \cup A \rightarrow C$ y $\alpha \cup B \rightarrow C$
(I.v.2) $B \rightarrow (A \vee B)$	(E.⊃.) $A,(A \supset B) \rightarrow B$
(I.⊃) $\alpha \rightarrow (A \supset B)$ si $\alpha \cup A \rightarrow B$	
(I.¬) $\alpha \rightarrow \neg A$ si $\alpha \cup A \rightarrow \bot$	(E.¬) $\alpha \rightarrow A$ si $\alpha \cup \neg A \rightarrow \bot$

Claramente cada uno de estos postulados puede reformularse con funciones de consecuencia al estilo de Tarski.

Igual procedimiento puede seguirse para las distintas lógicas: las lógicas modales, las de los llamados condicionales contrafácticos, las de los condicionales derrotables *(defeasible)*, los condicionales relevantes, las conjunciones asimétricas (temporales), la lógica intuicionista, etc., ya que se trata de un procedimiento general no circunscripto a ningún signo lógico ni sistema lógico en particular.

En la presentación de los axiomas anteriores se ha seguido una importante idea de Gentzen: la de dividir los principios que caracterizan cada signo lógico en dos categorías: los de Introducción y los de Eliminación. Los primeros regulan la figuración del signo en la conclusión de una inferencia (indican cómo introducir el signo en la conclusión en una derivación deductiva). Las segundas regulan la figuración del signo entre las premisas (indican cómo eliminar el signo al pasar de una premisa en que figura el signo a una conclusión).

Esta idea de Gentzen responde a una concepción acerca del significado de los signos lingüísticos característica del segundo período de la filosofía de Wittgenstein, según la cual el significado de un signo está determinado por las reglas que fijan su uso en cada contexto. La relación de consecuencia configura el contexto en el que la lógica se desarrolla, luego, el significado de un signo lógico se determina indicando cómo usarlo en las premisas y en la conclusión de la relación de consecuencia (los dos únicos lugares de esa relación). Esto es lo que se consigue al especificar las reglas de introducción y de eliminación.

El significado que el procedimiento anterior determina para un signo lógico, no es, por cierto, el significado semántico (del enfoque semántico), ya que él no depende de las correlaciones referenciales entre lenguaje y realidad efectuadas, en un enfoque semántico, por las funciones de interpretación, ni de la noción semántica de verdad. Tal significado se configura por medio de reglas sintácticas de inferencia, ya que esto es lo que en definitiva representan los axiomas anteriores de introducción y eliminación. De modo que es apropiado decir que en un enfoque abstracto los axiomas referidos a los signos lógicos determinan su *significado sintáctico*[13].

Así, en lo que hace a la especificación del significado de los signos lógicos e indirectamente a la identificación de cada lógica en particular, el enfoque abstracto comparte rasgos típicos del enfoque sintáctico.

Con frecuencia se piensa que para un enfoque sintáctico los signos del lenguaje objeto carecen por completo de significación. En rigor esto no es así, ya que los signos lógicos tienen un significado (sintáctico) que reciben de las reglas que fijan su uso en los contextos de consecuencia, si bien es cierto que los demás signos (los que no son lógicos) carecen totalmente de (o por lo menos no se considera, en un enfoque sintáctico o abstracto, su) significación. Por el contrario en un enfoque semántico todos los signos del lenguaje tienen (en cada modelo) significación. Los signos lógicos reciben su significación (semántica) a través de las cláusulas que determinan las condiciones de verdad de los enunciados en que ellos figuran (cláusulas que son comunes a todos los modelos de cada lógica). En este sentido, la pretensión (vinculada a veces a la tesis de la primacía de la semántica) de que sólo en un enfoque semántico los signos lógicos tienen signifido es un exceso equivocado.

Desde esta nueva perspectiva sólo el enfoque abstracto logra dar una caracterización general de la lógica, ya que sólo desde ese ángulo es posible explicar la razón por la cual lo que se define, tanto desde el enfoque semántico como desde el enfoque sintáctico, son efectivamente relaciones de consecuencia. La noción central de la lógica es, en este sentido, conceptualmente independiente de las características del método axiomático al que está anclado el enfoque sintáctico. La circunstancia de que la lógica sea un instrumento indispensable en la organización conceptual interna de cualquier disciplina a través de su organización deductiva (que justifica la peripatética concepción de la lógica como *organon* de todo conocimiento) y de que la lógica misma sea susceptible de una estructuración axiomática no significa que sólo a través de ese método ella tenga que ser identificada.

Lo anterior también implica que la lógica es de igual modo conceptualmente independiente de las nociones semánticas, y en particular de la noción de verdad. Nada de esto, por cierto, desmerece el valor de la

13. Para una defensa del enfoque abstracto como el expuesto en el texto, ver Belnap (1962) escrito en respuesta a la tesis de la primacía de la semántica sostenida por Prior (1960).

utilización del método axiomático en la lógica ni la importancia de las nociones semánticas en la apreciación de las calidades de cada lógica en particular.

Con frecuencia a lo largo de este artículo, siguiendo la forma más corriente de referirse al tema, hemos hablado de la lógica como si fuera una disciplina única. Sin embargo, para ser fieles al desarrollo histórico y a la situación actual, debe decirse que la expresión «Lógica» es un término genérico que se aplica a una pluralidad de disciplinas con características y aspiraciones diversas. Puede haber, y de hecho hay, muchas lógicas diferentes[14]. Cada lógica (cada uno de los ejemplos de la expresión genérica Lógica) es identificada extensionalmente (como también lo hemos hecho en este artículo) con una única relación de consecuencia. También hemos visto que cada lógica (cada relación de consecuencia) puede ser intensionalmente identificada básicamente a través de dos procedimientos diferentes que requieren la satisfacción de propiedades distintas de cada relación de consecuencia. Así, en un enfoque sintáctico en el que, por ejemplo, se use el esquema de definición (Def. 3) tienen que satisfacerse las propiedades sintácticas que esta definición requiere respecto de un conjunto de enunciados de $\mathbb{L}$ identificados como axiomas y respecto de un conjunto precisamente identificado como reglas de inferencia primitivas. Por el contrario, en un enfoque semántico en el que, por ejemplo, se use el esquema de definición (Def. 2.2), no son las anteriores las propiedades que la relación de consecuencia tiene que satisfacer sino que debe satisfacer las que (Def. 2.2) exige respecto de los modelos semánticos correspondientes. Sin embargo, lo que siempre tiene que satisfacer cada relación para ser una relación de consecuencia lógica deductiva son las tres propiedades postuladas en el enfoque abstracto. En este sentido el enfoque abstracto y sólo él, permite una definición del término genérico Lógica, ya que en él se contemplan las propiedades que cada uno de los individuos del género (cada una de las relaciones de consecuencia deductiva) tiene que poseer para pertenecer al género Lógica. Una relación de consecuencia es una lógica deductiva si y sólo si es reflexiva, monótona y valida el principio de «corte».

El permitir una definición general de la lógica es la virtud principal del enfoque abstracto. Sin embargo, no debemos olvidar que, completado con el sentido sintáctico de los signos lógicos anteriormente mencionado, el enfoque abstracto da respuesta al dilema de Jörgensen al explicar cómo son posibles lógicas referidas a enunciados que carecen de valor de verdad, superando así la dificultad indicada en VIII a la tesis de la primacía de la semántica.

14. Sobre el importante tema de si las diferentes lógicas son rivales entre si o si son complementarias y pueden ser integradas en única lógica omnicomprensiva, sólo se harán algunos breves comentarios en el apartado X.

X. LA JUSTIFICACIÓN INTUITIVA

En ciertos casos las reglas que determinan el significado sintáctico de los signos lógicos tienen, desde el punto de vista intuitivo, exactamente el mismo contenido conceptual que las cláusulas para la noción semántica de verdad que fijan el significado semántico de tales signos. Tal es claramente el caso cuando se comparan las reglas contenidas en los axiomas de introducción y de eliminación para la forma de conjunción considerada «∧» con la cláusula (iii) por la que se determina el significado semántico de dicha conjunción. En efecto, tanto unas como otras intuitivamente indican la identidad conceptual de una conjunción con el par de sus conyuntos.

Lo anterior no sucede (ni se pretende que suceda) con todas las maneras usadas para fijar el significado sintáctico y semántico de los diferentes signos lógicos que interese tematizar. Es un lugar común a todos los enfoques de la lógica que la apreciación intuitiva requiere considerar la totalidad de los principios que cada lógica convalida para cada uno de los signos lógicos en ella incluidos.

Así, por ejemplo, las llamadas «paradojas de la implicación material»:

(P.⊃.1) $A \rightarrow (B \supset A)$
(P.⊃.2) $\neg A \rightarrow (A \supset B)$

(que la implicación material comparte con la implicación intuicionista) muestran un desacuerdo conceptual entre, por un lado, los significados semánticos y sintácticos de las nociones de implicación de la lógica clásica y de la intuicionista, y por otro, los significados paradigmáticos de las construcciones condicionales de los lenguajes corrientes en uso.

La no satisfacción de la segunda (P.⊃.2) es la motivación principal que ha llevado a la construcción de las lógicas de los condicionales contrafácticos. Un condicional contrafáctico es aquél cuyo antecedente es de hecho falso. Aceptar (P.⊃.2) para los condicionales contrafácticos implica comprometerse a aceptar que todo condicional contrafáctico es siempre verdadero. Podemos creer que si Aristóteles hubiera conocido la lógica clásica contemporánea la habría aceptado, o podemos creer que si Aristóteles hubiera conocido tal lógica no la habría aceptado, lo que sí resulta completamente desquiciado es creer que hay alguna razón conceptual (vinculada a la noción corriente de condicionalidad) para tener que aceptar los dos condicionales anteriores por el sólo hecho de que Aristóteles, dada la época en que le tocó vivir, nunca conoció la lógica contemporánea.

Dar la satisfacción a estos desacuerdos intuitivos entre las lógicas conocidas y nuestra manera corriente de conceptualizar la realidad a través de los lenguajes que a diario usamos es la motivación explícita que subyace a la creación de las distintas lógicas para los condicionales contrafácticos. Además, es el tipo de motivación que justifica la creación de una gran mayoría del enorme espectro de lógicas que pueblan la lógica

contemporánea. Sin embargo, es dable reconocer que muchas lógicas han surgido motivadas sólo por consideraciones formales, como las de indagar qué sucede si se generaliza, se restringe o de alguna forma se modifica alguno o algunos principios de una lógica ya conocida.

A esta altura se plantean varios interrogantes filosóficamente relevantes:

1) ¿Por qué no hay una única lógica, por lo menos para cada uno de los términos sincategoremáticos del lenguaje?

2) Las fallas denunciadas como paradojas intuitivas ¿no son un síntoma de que se ha deslizado algo falso en las lógicas que las padecen?

3) La buscada correlación entre las lógicas desarrolladas con relación a lenguajes artificiales y el alcance conceptual de sus pretendidos correlatos de los lenguajes corrientes en uso, que por cierto es el instrumento con que pensamos a diario ¿no muestran lo justificado de la pretensión descriptiva de la antigua concepción de la lógica como la ciencia de nuestro modo de pensar (de las leyes del pensamiento)?

Es innegable que los lenguajes artificiales con relación a los cuales se identifican cada una de las distintas lógicas surgieron con el propósito de suministrar reconstrucciones racionales de conceptos que encuentran su expresión natural en los lenguajes corrientemente usados. También es verdad que no hay lógica que no esté afectada por alguna disonancia intuitiva. Tales discrepancias pueden detectarse tanto en sus presentaciones sintácticas como en sus presentaciones semánticas. Este aspecto es importante, porque muestra que tanto el significado sintáctico como el significado semántico de los signos de un lenguaje artificial son perfectamente inteligibles en sí mismos, ya que en caso contrario no podrían compararse con la significación del lenguaje natural y no existirían las paradojas intuitivas.

Lo anterior, sin embargo, no implica que la existencia de discrepancias intuitivas signifique que se haya deslizado alguna falsedad en la lógica en que ellas se producen. Así la conjunción clásica «$\wedge$», considerada usualmente como el signo lógico de máxima semejanza con sus correlatos del lenguaje corriente, tiene discrepancias intuitivas, ya que, tanto por su significado sintáctico como por el semántico, resulta que es una operación conmutativa, en el sentido de que el orden de los conyuntos es irrelevante: $(A \wedge B)$ significa lo mismo que $(B \wedge A)$. No obstante hay típicas construcciones conjuntivas del lenguaje corriente que carecen de esta propiedad: no es equivalente «se casaron y tuvieron un hijo» a «tuvieron un hijo y se casaron».

De ello no debe inferirse que haya algún error en las leyes y reglas de la conjunción clásica, ya que ellas están plenamente justificadas por el significado (sintáctico y semántico) atribuido a esa forma de conjunción. Lo que la discrepancia intuitiva muestra es que la conjunción clásica no reconstruye todas las construcciones conjuntivas corrientes. La discrepancia sólo muestra la necesidad de restringir el *ámbito de aplica-*

bilidad (con relación a las construcciones corrientes) de esa forma de conjunción. El propósito reconstructivo se vería frustrado sólo cuando tal ámbito fuera vacío. Pero aun en ese caso no significa la inclusión de falsedad alguna en la lógica clásica de la conjunción, ya que es sólo su valor práctico de utilidad el que se encontraría cuestionado, valor que podría reivindicarse con razones diferentes.

Desde la anterior perspectiva la coexistencia de una pluralidad de lógicas, cada una de ellas plenamente justificadas en sí mismas, no debe ser motivo de extrañeza. Es más, los principios de cada lógica pueden ser vistos como analíticos en el sentido en que ellos se justifican apoyándose únicamente en el significado (sintáctico o semántico según cuál sea la naturaleza del enfoque con que se identifica la significatividad de los signos lógicos) de sus expresiones constituyentes. La tesis expuesta corresponde sustancialmente a la de Quine (Quine, 1970) frente a la cuestión de las lógicas alternativas y rivales que él sintetiza en el dictum: cambio de lógica implica cambio de tema. No hay ni puede haber rivalidad entre dos lógicas diferentes porque un cambio en los principios supone un cambio en el significado de los signos lógicos que en ellos figuran.

Así, por ejemplo, la lógica de la implicación intuicionista está caracterizada por la adopción de los dos axiomas (I. $\supset$) y (E. $\supset$) como criterio para el significado (sintáctico) del signo de implicación. La implicación material se constituye cuando además de los anteriores se postula el siguiente axioma de eliminación[15]:

(E. $\supset$.1) $((A \supset B) \supset A) \rightarrow A$ (Principio de Peirce)

Naturalmente una caracterización semántica de la implicación intuicionista requiere una estructura semántica[16] distinta de la ofrecida anteriormente para el condicional clásico y condiciones de verdad distintas a las incluidas en la cláusula (v).

Las dos lógicas son diferentes porque sus principios no son los mismos (en la intuicionista el principio de Peirce no vale), pero esto implica que el signo de implicación tiene en cada una de ellas un significado diferente.

Con este alcance dos lógicas diferentes no pueden ser rivales. Naturalmente, esto no excluye una necesaria rivalidad en el ámbito de su aplicación. Pero con ello nos vamos del área de la significatividad sintáctica y semántica para entrar en el de la justificación intuitiva en la comparación con las construcciones correspondientes del lenguaje corriente. Es sensato pensar que es precisamente en este área donde hay que buscar una de las fuentes, no por cierto la única, de justificación de toda lógica.

La justificación pragmática de una lógica por el ámbito del lenguaje corriente que logra reconstruir es quizás el grano de verdad contenido

15. La circunstancia de que en la presentación precedente de la lógica clásica no se incluyó al principio de Peirce es porque en presencia de (I. $\neg$) y (E. $\neg$), (E. $\supset$. 1) se deriva de (I. $\supset$) y (E. $\supset$).

16. La semántica estándar para la lógica intuicionista, al igual que la semántica de las lógicas modales, recurre a la noción de «mundo posible».

en la concepción tradicional de la lógica como versando acerca de las «leyes del pensamiento».

Lo anterior no descalifica la pretensión de buscar reemplazar la pluralidad de lógicas por una única lógica general. La idea sería la siguiente: incorporar en un único sistema de lógica todos los signos lógicos que tengan un ámbito de aplicación no vacío, precisando además cuál es el ámbito de aplicación de cada uno de los signos lógicos con significado (semántico y sintáctico) diferente.

Si tal proyecto es o no, en definitiva, realizable, no podemos responder. El hecho es que hasta ahora no es mucho lo que se ha hecho en esa dirección. No obstante, hay que tomar en cuenta la siguiente advertencia limitativa.

No es sensato esperar entre la lógica reconstruida en un lenguaje artificial y la «lógica del lenguaje natural» una suerte de correlación como la requerida en (Corr. 1) entre los enfoques sintáctico y semánticos, porque en un sentido importante no hay una lógica coherente en el lenguaje natural. El lenguaje corriente no sólo está plagado de ambigüedades, vaguedades y toda suerte de imprecisiones significativas que justifican apartarse de él en los procesos de reconstrucción racional, sino que acumula en su seno intuiciones incompatibles que no pueden superarse más que reformándolo, abandonando intuiciones que pueden ser muy sólidas.

La siguiente situación es un ejemplo de la dificultad anterior.

Pocas cosas son más intuitivas que la necesidad de rechazar el principio de la lógica clásica (compartido con casi todas las lógicas salvo las lógicas relevantes y las paraconsistentes) de que de dos enunciados contradictorios (y de todo conjunto inconsistente) todo enunciado es consecuencia deductiva. Lo que se rechaza es el principio de Duns Escoto:

(DE) $A, \neg A \rightarrow B$

Igualmente es muy difícil no admitir la enorme intuitividad de la regla de Introducción de la Disyunción (I.v.1), el principio de monotonía (→.3), el principio del Silogismo Disyuntivo:

(SD) $\neg A, (A \vee B) \rightarrow B$

y de la suerte de transitividad de la noción de consecuencia representada por la regla de Corte (→.2).

Lamentablemente todas estas fortísimas intuiciones no pueden convivir coherentemente ya que la aceptación de los cuatro últimos principios compromete a la aceptación del principio de Duns Escoto. La siguiente es una derivación de (DE) a partir de los otros cuatro principios indicados:

(1) $A, \neg A \rightarrow (A \vee B)$ (De (I.v.1) por (→.3))
(2) $A, \neg A, (A \vee B) \rightarrow B$ (De (SD) por (→.3))
(3) $A, \neg A \rightarrow B$ (De (1) y (2) por (→.2))

La más leve aspiración de coherencia nos competen a rechazar alguna de estas sólidas intuiciones, pero por otro lado no hay base intuitiva para el sacrificio de ninguna.

La conclusión de esta paradoja intuitiva es que cualquiera que sea la lógica que terminemos privilegiando, ella tendrá que apartarse de las intuiciones básicas incorporadas al esquema de conceptos de los lenguajes corrientes. Esto implica abandonar una idea reconstructivista con pretensiones de resultados unívocos.

Mantener a toda costa el rechazo del principio de Duns Escoto es la motivación subyacente a la construcción de las llamadas Lógicas Relevantes, y es también una de las motivaciones más importantes de las Lógicas Paraconsistentes. En verdad en ellas se reconstruyen intuiciones fundamentales, pero queda abierta la pregunta pragmática de si no es más conveniente dejar tales intuiciones a un lado ya que, en definitiva, no se puede dar satisfacción a todas las intuiciones corrientes.

Estas últimas reflexiones quizás recojan el grano de verdad contenido en la vieja concepción normativa de la lógica como disciplina acerca de cońo se debe pensar.

BIBLIOGRAFÍA

Belnap, N. D. (1962), «Tonk, Plonk, and Plink»: *Analysis, 22,* incluido en P. F. Strawson (ed.), *Philosophical Logic,* OUP, Oxford, 1967.

Carnap, R. (1937), *The Logical Sintax of Language,* Routledge&Kegan Paul, London.

Carnap, R. (1942), *Introduction to Semantics,* Harvard Univ. Press.

Carnap, R. (1943), *Formalization of Logic,* Harvard Univ. Press.

Carnap, R. (1955), «Foundations of Logic and Mathematics», en O. Neurath, R. Carnap, y Ch. Morris (eds.), *International Encyclopedia of Unified Sciences,* vol. I, n.º 3, The University of Chicago Press, Chicago.

Gentzen, G. (1934), «Untersuchungen über das Logische Schliessen»: *Mathematische Zeitschrift, 39,* 176-210, 403-431.

Prior, A. N. (1960), «The Runabout Inference-Ticket»: *Analysis, 21,* incluido en P. F. Strawson (ed.), 1967.

Quine, W. v. O. (1970), *Philosophy of Logic,* Prentice Hall.

Tarski, A. (1930a), «On Some Fundamental Concepts of Mathematics», incluido en *Logic, Semantics, Metamathematics,* Clarendon Press, Oxford, 1956.

Tarski, A. (1930b), «Fundamental Concepts of the Methodology of the Deductive Logic», incluido en Tarski, 1956.

Tarski, A. (1935), «The Concept of Truth in Formalized Languages», incluido en Tarski, 1956.

Tarski, A. (1936), «On the Concept of Logical Consecuence», incluido en Tarski, 1956.

Wright, G. H. von (1957), *Logical Studies,* Routledge&Kegan Paul, London.

HISTORIA DE LA LÓGICA

José A. Robles García

En las siguientes páginas se presenta un sucinto panorama histórico de la lógica dividido en las siguientes secciones: I. Lógica griega; II. Lógica medieval; III. La lógica antes de Frege; IV. La lógica de Frege; V. Después de Frege.

Ciertamente muchas cosas quedarán fuera de este panorama y el detalle de algunos de los diferentes temas que aquí presentaré o a los que sólo aludiré, lo encontrará el lector en los diversos artículos de esta enciclopedia dedicados precisamente a esa tarea.

El tema de la lógica medieval, aun cuando *per se* es de gran importancia por la variedad y riqueza de tratamiento de diversos tipos de inferencias, me veo precisado a bosquejarlo tan solo, tomando tres temas centrales, pocos ejemplos y menos autores y matices.

I. LÓGICA GRIEGA

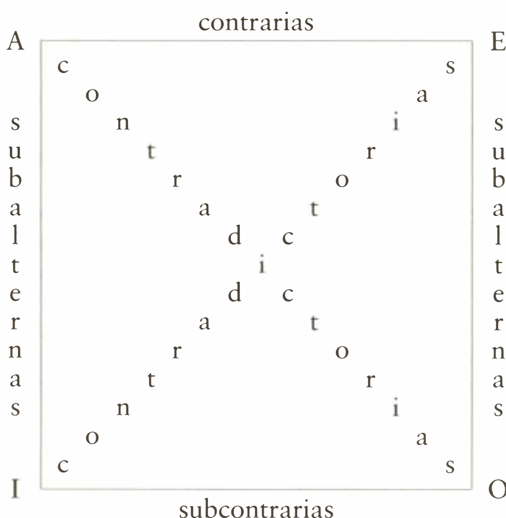
1. *Aristóteles*

Con el trabajo de Aristóteles (384-322 a.C.) surge la lógica en el mundo. Él desarrolla la teoría del silogismo, a la que se alude cuando se habla de la lógica aristotélica como un antecedente remoto de la lógica contemporánea.

En el despliegue de la parte central de su teoría, Aristóteles sólo considera cuatro tipos diferentes de enunciados o proposiciones a partir de los cuales formula sus propuestas de argumentación válida. Los cuatro enunciados (o, mejor, *formas enunciativas,* esto es, expresiones en las que figuran variables y que se convierten en enunciados una vez que estas variables se sustituyen por las expresiones adecuadas correspondientes) en cuestión son el universal afirmativo, 'Todo S es P' (A), el universal

negativo, 'Ningún S es P' (E), el particular afirmativo, 'Algún S es P' (I) y el particular negativo 'Algún S no es P' (O). En donde las letras 'S' y 'P' son variables predicativas que toman como valores sustantivos generales, de tal manera que una forma enunciativa del tipo (I), 'Algún S es P', se convierte en el enunciado 'Algún *hombre* es *mortal*' al sustituir 'S' y 'P' por 'hombre' y por 'mortal', respectivamente. Finalmente es preciso señalar que Aristóteles también consideró otra clase de enunciados, los individuales, como 'Sócrates es mortal', que comentaristas posteriores asimilaron a los enunciados universales afirmativos, creando bastante confusión, pues la predicación tiene características diferentes en ambos casos.

La manera gráfica, postaristotélica, de representar las relaciones lógicas entre los enunciados (formas enunciativas) categóricos aristotélicas (A, E, I, O), se conoce con el nombre de *cuadrado de oposición* y es el siguiente:



Las relaciones lógicas que se dan entre estos enunciados son: *los contrarios* (A, E; esto es, los enunciados universales), pueden ser *ambos falsos,* pero no *ambos verdaderos; los subcontrarios* (I, O; esto es, los enunciados particulares), en cambio, pueden ser *ambos verdaderos,* pero no *ambos falsos;* por otra parte, con respecto a *la subalternación,* de la verdad de cualquiera de los contrarios (A, E), se sigue la verdad del subcontrario correspondiente (I, O) y de la falsedad de cualquiera de los subalternos (I, O), se sigue la falsedad del contrario correspondiente (A, E). Finalmente, *los enunciados contradictorios* tienen siempre valores veritativos opuestos: *si uno de ellos es verdadero, el otro es falso y a la inversa.*

La silogística aristotélica forma parte de la que hoy se considera la teoría general de la inferencia deductiva. Conforme a ésta, se define lo que es un argumento deductivo *válido* (y aquí tendremos que apelar a

las formas enunciativas, sin entrar en demasiados detalles) como un conjunto de enunciados, e_1, e_2,..., e_n, tales que éstos *provienen de formas enunciativas* tales que es *imposible* que haya una sustitución (de las variables por predicados) en la que los enunciados resultantes, e_j ($1 \leqslant j < n$), sean todos verdaderos y e_n sea falso. A los enunciados e_j se les denomina *premisas* del argumento y al enunciado e_n se le llama *conclusión*. En la silogística aristotélica se estudian argumentos formados por sólo dos premisas y conclusión, donde estos tres enunciados (formas enunciativas) serán, todos, sólo de alguna de las formas A, E, I, O arriba mencionadas.

Aristóteles desarrolla la teoría del silogismo considerando *todas* las formas válidas posibles de inferencia dentro de este esquematismo lógico. Conforme a él, son válidas las inferencias que, de un enunciado universal como premisa van a un enunciado particular como conclusión, es decir, los siguientes esquemas (formas de argumento) muestran formas válidas de inferencia:

Todo S es P	Ningún S es P
∴ Algún S es P	∴ Algún S no es P

Estos esquemas (que corresponden a la relación de *subalternación; cf. supra,* p. 50), aun cuando tienen una apariencia intuitiva correcta, fueron puestos en cuestión, como veremos (*cf. infra,* pp. 63 ss.), por los lógicos contemporáneos.

Los silogismos aristotélicos, según lo señalé, son esquemas de argumentación compuestos de dos premisas y una conclusión. En los silogismos, para ser tales, deben de figurar, entre premisas y conclusión, *exactamente* tres términos, es decir, sólo tres expresiones diferentes de las que forman los sujetos y los predicados de los tres enunciados. Un ejemplo de silogismo esquema es el siguiente:

premisa menor	Todo B es C
premisa mayor	Todo C es D
conclusión	∴ Todo B es D

En donde los tres términos diferentes están representados por las tres letras 'B', 'C' y 'D'. A partir de esta composición los términos reciben los siguientes nombres: 'D', *el predicado de la conclusión,* se denomina *término mayor;* 'B', *el sujeto de la conclusión,* se denomina *término menor* y, finalmente, 'C', *el término que figura sólo en las dos premisas,* se denomina *término medio.* A partir de estas denominaciones, las premisas reciben los nombres de, *premisa mayor,* aquella en la que figura el término mayor y *premisa menor,* aquella en la que figura el término menor, sin importar el orden en el que tales premisas se encuentren colocadas.

Tras las observaciones anteriores, podemos entender la agrupación que hace Aristóteles de los silogismos en tres figuras bajo las que caen diversos modos. Las figuras dependen de la colocación de los términos en las premisas y los modos dependen del tipo de enunciados que for-

men el silogismo. De acuerdo con esto, las figuras del silogismo aristotélico son:

1ª figura	2ª figura	3ª figura	
C-D	D-C	C-D	mayor
B-C	B-C	C-B	menor
B-D	B-D	B-D	conclusión

Y los modos en cada una de las figuras anteriores, son:

1ª figura	2ª figura	3ª figura
A, A/A	E, A/E	A, A/I
E, A/E	A, E/E	A, I/I
A, I/I	E, I/O	E, A/O
E, I/O	A, O/O	E, I/O
		I, A/I
		O, A/O

Lo que muestra el esquema anterior es el tipo de premisas y de conclusión que pueden figurar en esquemas válidos (*cf. supra,* p. 51) de silogismo en cada una de las tres figuras. Así, el caso A, A/A de la primera figura señala que en esa figura es válido un silogismo cuyas dos premisas y conclusión sean enunciados universales afirmativos (A). En la segunda figura la conclusión siempre es negativa (E u O) y, *por esto,* una de las premisas debe se negativa. En la tercera figura todas las conclusiones son particulares (I u O).

Las reglas de construcción de los silogismos determinan cuáles son las formas válidas posibles dentro de cada una de las figuras.

A la muerte de Aristóteles se añade, a las tres figuras aristotélicas, una cuarta figura (atribuida falsamente —según los Kneale— a Galeno, siglo II de nuestra era), con seis modos válidos. Esta cuarta figura invierte la estructura de las premisas de la 1ª figura:

	4ª figura	modos	
premisa mayor	D-C	A, A/I	A, E/E
premisa menor	C-B	I, A/I	E, A/O
conclusión	∴ B-D	E, I/O	A, E/O

2. *Megáricos y estoicos*

En el caso de los megáricos y de los estoicos poco podemos decir, ya que hay escaso material conservado acerca de su trabajo. Los megáricos, señalan los Kneale, hicieron tres aportaciones a la lógica en lo relativo a las paradojas, a una destacada revisión de los conceptos modales y comenzaron un importante debate con relación a los enunciados condicionales. En lo que sigue algo se dirá acerca de la tercera propuesta y muy poco acerca de las dos primeras.

De las paradojas megárico-estoicas que hasta nosotros han llegado hay que subrayar la muy conocida del mentiroso, atribuida a Eubúlides: Si alguien dice '*estoy mintiendo*', ¿es verdadero o falso esto que dice? Aquí vale recordar mínimamente las alternativas: si lo que dice *es verdad,* entonces, *está mintiendo,* por tanto, lo que dice es falso, entonces *no está mintiendo;* por otra parte, *si lo que dice es falso,* entonces *no está mintiendo,* esto es, *está diciendo la verdad,* a saber, *está mintiendo.* La conclusión es, entonces, que si está mintiendo está diciendo la verdad y si está diciendo la verdad, entonces está mintiendo. Esto muestra que el sujeto, *a la vez,* miente y *no* miente. Pero los dos enunciados *no* pueden ser verdaderos dentro de una lógica bivaluada en la que vale una ley equivalente a la señalada para los enunciados contradictorios (*cf. supra,* pp. 50-51).

Con respecto a la naturaleza de los enunciados condicionales, los primeros en estudiarlos, de acuerdo a los Kneale, fueron Diodoro Crono y su discípulo Filón. Lo que nos dicen acerca de esto es que «Sexto Empírico, al reseñar la disputa sobre los condicionales, señala que Filón sostiene que un condicional correcto (*ὑγιὲς συνημμένον*) es uno que no comienza con una verdad y concluye con una falsedad; pero Diodoro dice que un condicional así es uno que no comienza ni puede comenzar con una verdad y acabar con una falsedad...». En el primer caso, el de Filón, tenemos la que Russell denominó 'implicación material'; conforme a la caracterización de Filón, un condicional como «si ahora es de día entonces $2 + 2 = 5$» será verdadero si se dice de noche, en tanto que nunca será verdadero, de acuerdo a la caracterización de Diodoro, ya que puede comenzar con una verdad, si el condicional se dice de día, y concluirá con una falsedad. De acuerdo a la propuesta de Diodoro, entonces, un condicional será verdadero con sólo que su antecedente (o la parte que viene después de 'si' y antes de 'entonces') sea *siempre* falso o que su consecuente (o la parte que viene después de 'entonces') sea *siempre* verdadero; de esta manera, se cumple con la exigencia de que nunca se dé el caso de que el antecedente sea *verdadero* y el consecuente *falso.*

La implicación de Filón (implicación material) es la implicación de la lógica clásica contemporánea, en tanto que el último tipo de implicación, la de Diodoro, es la relación de implicación lógica o implicación estricta que adoptó C. I. Lewis en su lógica modal.

II. LÓGICA MEDIEVAL

Los lógicos medievales no llegan a formular una teoría lógica tan plenamente formalizada como la que tenemos hoy en día por la razón de que su interés se centraba en estudiar y formular las leyes lógicas de una lengua natural, el latín, a diferencia de la práctica de los lógicos contemporáneos, cuyo interés es, más bien, el estudio y *la construcción* de lenguajes simbólicos (artificiales), que tengan ciertas propiedades que se consideran útiles teniendo en cuenta ciertos propósitos a la vista.

Sin embargo, un rasgo básico que establece una fuerte liga entre los lógicos medievales y el trabajo contemporáneo en esta disciplina, es la clara conciencia que tenían los primeros de que la misma es un estudio de estructuras formales y que precisamente era por medio de la forma de las proposiciones como se podrían evaluar los argumentos y determinar los casos de consecuencias lógicas correctas.

Para poner lo anterior de relieve sigo, en parte, a Philoteus Böhner y señalo las que él considera las principales aportaciones de esta época. Los encabezados bajo los que las reúne, son: 'Términos sincategoremáticos', 'Teoría de la suposición' y 'Teoría de las consecuencias'.

1. *Términos sincategoremáticos*

Sobre los *términos sincategoremáticos* o palabras cosignificantes, Böhner señala que hay una fuerte relación entre el uso escolástico y el estoico de las mismas. En ambos casos se usa la misma palabra y se le da el mismo significado. Bochenski también encuentra una fuerte relación con los estoicos y señala, por su parte (1961, 189), que la teoría de las consecuencias es «esencialmente un avance de la lógica proposicional estoica», aun cuando, quizás, no hubo ninguna influencia estoica directa en la construcción medieval.

Los términos sincategoremáticos se contrastan con los categoremáticos; éstos, según lo señala Alberto Magno, son los «que, tomados significativamente, pueden ser sujetos o predicados —o parte del sujeto o parte del predicado distribuido— de una proposición categórica; por ejemplo, los términos 'hombre', 'animal', 'piedra', se llaman categoremáticos porque tienen una significación definida y cierta» (*Perutilis logica,* § 44). Los términos sincategoremáticos, en cambio, son los que *no* pueden ser sujetos ni predicados de una proposición (a menos que se tomen materialmente, como en «'Y' es una conjunción») y, en el caso preciso de la lógica, esos términos son los «signos universales o particulares» (como los llama Alberto: *ibid.*), que son nuestros cuantificadores, así como las conectivas lógicas: negación, conjunción, disyunción, etc. Lo que Böhner señala acerca de estos términos es que la importancia que los escolásticos les dieron señala con claridad que tenían muy en cuenta el carácter *formal* de sus investigaciones. Los términos sincategoremáticos que aquí hemos señalado son los pertinentes para el estudio de la lógica, ya que influyen directamente en la verdad o en la falsedad de las proposiciones.

Para mostrar cómo entendían los autores medievales la función de los términos sincategoremáticos, vuelvo a citar a Alberto de Sajonia quien, al responder la objeción de que, aparentemente, «y» puede formar parte del sujeto de una proposición, como, por ejemplo, en 'Sócrates y Platón corren', parece ser que el sujeto es «Sócrates y Platón», Alberto dice,

> ... en la proposición 'Sócrates y Platón corren', «y» no es parte del sujeto, sino que solamente son sujetos el término 'Sócrates' y el término 'Platón'; esto es claro, pues

la contradictoria es 'Sócrates o Platón no corren'; pero ambas proposiciones no serían contradictorias si «y» fuese parte del sujeto de la primera, pues entonces las dos proposiciones no tendrían el mismo sujeto.

(*ibid.* § 52)

De las consideraciones anteriores, Alberto precisa los sentidos de *consecuencias material y formal* y, según lo señala acertadamente Böhner, el uso que se les da a los términos categoremáticos escolásticos es como el de nuestras variables predicativas y los términos sincategoremáticos se presentan como nuestras constantes lógicas. En otro pasaje de su *Perutilis logica,* Alberto nos dice:

975. De las consecuencias, una es formal y otra material...

976. ... y tal como hablo aquí de formas y materia, se entiende por materia de la proposición o de la consecuencia, los términos puramente categoremáticos —como son los sujetos y los predicados—, prescindiendo de los sincategoremas que les acompañan y por los que tales se coordinan o distribuyen, al ser llevados a un determinado modo de suposición.

977. Y a la forma se dice que pertenece todo lo demás; de modo que la cópula, tanto de la categórica como de la hipotética, pertenece a la forma de la proposición. Del mismo modo, las negaciones y los signos y el orden mutuo de éstos y los modos de significar pertinentes a la cantidad de la proposición categórica, como la discreción, comunidad, etc...

(§ 975-7)

Al considerar la teoría de las consecuencias volveremos sobre el tema que aquí aparece apuntado acerca, justamente, de las consecuencias formales, que según dije al iniciar esta sección, son el núcleo del estudio de la lógica.

2. *Teoría de la suposición*

Bajo el nombre de *proprietates terminorum,* se estudiaron, principalmente en la baja Edad Media, la *significatio,* la *copulatio,* la *appelatio* y la *suppositio.* La teoría de las *proprietates terminorum* creció en complejidad a medida que diversos autores daban diferentes formulaciones de las mismas e introducían matices y distinciones varias en ellas. Nuevamente aquí se hace palpable la diferencia, ya señalada, en la manera de tratar la lógica por parte de los escolásticos y de los autores modernos. Los primeros ofrecen análisis del latín (un lenguaje natural), a diferencia de los segundos que proponen sus análisis de un lenguaje artificial.

A pesar de las diferencias en el tratamiento, Böhner (pp. 29-30), encuentra similitudes entre la teoría de la *suppositio* medieval y el cálculo funcional contemporáneo. En cambio, acerca de la misma *suppositio,* Bochenski (1961, 162-3) nos dice que la doctrina fue una de las creaciones más originales de los escolásticos, pero desconocida tanto para la lógica antigua como para la moderna.

Aquí sólo menciono la doctrina y algunas de los importantes autores que escribieron sobre ella: Guillermo de Shyreswood, Tomás de Aquino, Vicente Ferrer, Walter Burleigh, Guillermo de Ockham, etc.

3. *Teoría de las consecuencias*

Una doctrina en la que puedo detenerme más, ligada con mayor claridad a nuestra lógica formal contemporánea, es la teoría de las consecuencias. El origen de una doctrina clara Bochenski (1961, 199-200) lo atribuye a Buridán, a Ockham e incluso a Pedro Hispano. La doctrina, fuertemente influenciada por Buridán, aparece con precisión en la *Perutilis logica* de Alberto Magno, la que Filoteo Böhner considera una piedra miliar en la teoría de la consecuencia y, afirma, «estamos firmemente convencidos de que, en muchos respectos, es superior a la *Summa Logicae* de Ockham» (Böhner, 70). Alberto, en el libro mencionado, luego de considerar diversas definiciones de antecedente y consecuente, formula la siguiente caracterización de estas expresiones en una relación de consecuencia:

> ... una proposición es antecedente de otra si se relaciona de tal manera con ella que es imposible que las cosas sean como, del modo que sea, las significa la primera —siempre que se mantenga fijo el uso de los términos—, sin que sean como las significa la otra.
>
> (IV, i, § 962)
>
> (versión modificada, siguiendo las sugerencias de Böhner y de Bochenski, de la traducción de Ángel Muñoz).

Lo que propone Alberto, entonces, de manera muy similar a la de Diodoro (*cf. supra,* p. 53), es afirmar que una proposición es antecedente de otra (se refiere al objeto que sea y siempre que los términos se apliquen de igual manera) si no es posible que la primera sea verdadera y no lo sea la segunda. (Véase la justificación que da Böhner de esta lectura en Böhner, 71-72.) Vale la pena añadir aquí, que esta lectura de Alberto *no* convierte en modal la proposición de consecuencia, pues la imposibilidad (necesidad) no se predica de la proposición misma, sino del valor de verdad de tal conexión (*cf. ibid.).*

Tal como Böhner analiza el condicional de Alberto, llega a la conclusión de que, para éste, se trata de una implicación formal o necesaria, no del tipo de implicación material, a la manera de Filón.

Más adelante, en la misma *Perutilis logica,* Alberto hace una distinción entre implicaciones o consecuencias *formal* y *material;* así, en *ibid.* § 975, señala (ya citado, en parte en *supra,* pp. 54-55):

> De las consecuencias, una es formal y otra material. Consecuencia formal se llama a toda proposición semejante en la forma a la que, si se formara, fuese buena consecuencia, como aquí: 'B es A; luego, lo que es A es B'.

Y, en la siguiente sección, señala:

> § 976. Consecuencia material es aquella tal que no toda proposición semejante a ella en la forma es buena consecuencia [...] y [...] se entiende por materia de la proposición o de la consecuencia, los términos puramente categoremáticos —como son los sujetos y los predicados—, prescindiendo de los sincategoremas que los acompañan...

Con la cita anterior cierro esta breve incursión por el mundo de la lógica medieval, en el que pudimos atisbar intuiciones muy lúcidas que, en muchos casos, tuvieron muy amplio desarrollo posterior gracias a la creación de lenguajes simbólicos adecuados que fueron vehículos más ágiles para el manejo lógico que los lenguajes naturales en los que se formulaba anteriormente la argumentación.

III. LA LÓGICA ANTES DE FREGE

1. *Leibniz y su idea de un lenguaje universal*

Tras la propuesta de Ramón Llull (1235-1315), en su *Ars Magna,* de formular un lenguaje universal de razonamiento, fundado en el supuesto de que todo el conocimiento no es sino un complejo que se forma a partir de la unión de ideas básicas, simples, muchos intentos se hicieron por formular un lenguaje de esta naturaleza. En el siglo XVII diversos pensadores hacen propuestas más claras y precisas que las de Lulio; entre ellos, Descartes formuló una propuesta a este respecto (carta a Mersenne del 20 de noviembre de 1629), en la que alude al orden numérico y a la formación de nuestros pensamientos a partir de pensamientos simples. En Inglaterra también se hicieron propuestas (John Wilkins [1614-1672] y George Dalgarno [1626-1687]) de construir un lenguaje en base a principios simples y con una gramática regular. Lo que esto daría como resultado sería facilitar la comunicación y hacer, por esto, más rápida la difusión de las ideas.

Vale la pena destacar aquí la aparición, en 1662, de uno de los libros más influyentes de la época, *La logique ou l'art de penser* (conocida como la *Lógica de Port Royal*), de los pensadores jansenistas de Port Royal, Antoine Arnauld y Pierre Nicole. Este libro se siguió imprimiendo hasta el siglo XIX. Los temas que trata, sin embargo, no son lo que hoy en día reconoceríamos como temas de lógica y, según lo señalan los Kneale, «... es la fuente de la mala costumbre de confundir la lógica con la epistemología». Sin embargo, hay que señalar que, entre otras cosas, en el libro se hace la distinción entre extensión y comprensión de un término.

Uno de los grandes filósofos de la época moderna que más se interesó por la lógica y por la creación de un lenguaje simple de razonamiento, fue G. W. Leibniz (1646-1716) quien, a los 19 años, llegó a acariciar

el proyecto de construir una *lingua philosophica* o *characteristica universalis,* esto es, un lenguaje que reflejara la estructura del pensamiento y que, por su medio, se pudiera llevar a cabo un cálculo que permitiera decidir todas las cuestiones relacionadas de consistencia y consecuencia. La propuesta de Leibniz, según se ha señalado aquí, no fue la primera pero sí más elaborada que las que se hicieron previamente. Leibniz formula su proyecto en su texto *De Arte Combinatoria* de 1666. La manera como él vislumbraba este nuevo lenguaje era en términos de una analogía con la construcción de los enteros: así como todos los números enteros o bien son primos o bien pueden obtenerse como productos de primos (2, 3, 5, 7, 11, 13, 17, 23, etc. o bien $4=2^2$, $6=2\times3$, $8=2^3$, $9=3^2$, $10=2\times5$, $12=2^2\times3$, etc.), así, en este nuevo lenguaje leibniziano, se podrían expresar las ideas simples (el equivalente de los números primos) y las ideas complejas (que serían un compuesto de ideas simples).

Leibniz intenta reflejar la complejidad de nuestro pensamiento en la simplicidad de la estructura matemática; el problema aquí es que, junto con la estructura, se requiere dar un análisis de los contenidos de los pensamientos y esto va más allá de lo que se puede hacer con sólo un análisis del lenguaje, en caso de que esto fuera todo lo que Leibniz deseara hacer.

Sin embargo, para que surgieran realmente cambios y se avanzara en lógica habría que esperar hasta el siglo XIX. Como es bien sabido, Kant, en el 'Prefacio' a la segunda edición (1787) de su *Crítica de la razón pura* (B viii), dejó sentado que la lógica, desde Aristóteles, no había avanzado nada y, así, señala que '... tiene toda la apariencia de ser perfecta y estar completa'.

2. *Antecedentes matemático-geométricos de la lógica actual*

Para llegar a presenciar los cambios en la visión de la lógica que surgen en el siglo XIX, es preciso tener en cuenta los avances en la investigación en matemáticas que dan origen, entre otras cosas, al surgimiento del álgebra abstracta, de las geometrías no euclídeas y a la preocupación por determinar la consistencia de la matemática misma.

Con respecto al álgebra abstracta, los trabajos en el siglo XIX, de Peacock, Hamilton, Abel, Galois, Cayley, etc. muestran que las operaciones aritméticas, hasta entonces usadas con un solo significado, podían redefinirse según diversas necesidades; de esta manera, una operación como la multiplicación, $\times$, por ejemplo, podría no ser conmutativa en el caso de los cuaternios hamiltonianos o de los vectores.

Con respecto al surgimiento de las geometrías no euclídeas, el siglo XIX presenció el desenlace de la larga historia con respecto a si el postulado V de Euclides, el postulado de las paralelas, era o no independiente de los otros postulados de los *Elementos.* Decir que un enunciado *e,* es *independiente,* de otros enunciados $e_1, e_2, \ldots, e_n$, es decir que *e* no es una

conclusión deductiva de los enunciados e_j ($1 \leqslant j \leqslant n$) o bien que no es contradictorio añadir al conjunto de enunciados e_j *la negación* del enunciado *e*, $-e$. Los resultados que se obtienen en el siglo XIX son de geometrías en las que por un punto *p*, exterior a una recta *r*, puede trazarse o bien *más de una* recta, *r'*, paralela a *r* (Lobachevsky) o bien *ninguna* recta *r'* paralela a *r* (Riemann). Recordemos que, en la geometría de Euclides, por un punto *p*, exterior a una recta *r*, podía trazarse *exactamente* una recta *r'* paralela a *r*.

Los resultados anteriores, y algunos más, hacen que los matemáticos de la época se preocupen por la consistencia de su herramienta de trabajo. También en el siglo XIX se comienza a elaborar la axiomatización de los números reales y a adquirir conciencia de las relaciones entre los diversos tipos de números que hasta entonces se habían estado usando sin preocuparse por las posibles relaciones que entre ellos pudieran existir.

Lo que ahora es importante señalar, en base a lo que hasta aquí se ha dicho, es que todos los anteriores avances ayudaron a que los matemáticos tomaran conciencia de que podían modificar, negar o rechazar principios asumidos que sólo la costumbre había hecho que parecieran inamovibles. Los resultados que podían obtener serían no sólo consistentes, sino también interesantes, teniendo en cuenta las posibles aplicaciones de los nuevos sistemas recién formulados o bien incluso por sí mismos, por las relaciones que mostraban que se daban entre sus elementos. En el campo de la lógica los avances en álgebra influyen de manera importante la labor de George Boole.

3. *Boole y el álgebra de la lógica*

El trabajo de Boole tiene como antecedente inmediato la labor de De Morgan y de Hamilton con relación a los viejos enunciados aristotélicos A, E, I, O. Si en la tradición aristotélica anterior, el sujeto y el predicado de los enunciados se veían como signos de cualidades, De Morgan y Hamilton los ven como signos de *las cosas que tienen esas cualidades*. Por otra parte, en la tradición aristotélica, los enunciados afirmativos, A, I, se explicaban como relacionando el sujeto con sólo parte del predicado; así, 'Todo S es P' se entendía como afirmando que la cualidad de ser P era parte de la cualidad de ser S pero, además, no se agotaba P en ser S; en terminología tradicional, en los enunciados afirmativos no estaba distribuido el predicado. A diferencia de esto, en los enunciados negativos el predicado sí estaba distribuido. Hamilton, además de introducir una manera diferente de ver los términos de los enunciados, considera la posible cuantificación del predicado y, así, es posible tener dos enunciados de tipo A: 'Todos los S son todos los P' así como 'Todos los S son algunos P'. Con todos estos elementos a la mano, es posible dar una interpretación de los enunciados como afirmando relaciones entre clases de objetos y, entonces, formular las relaciones entre éstas en términos de un álgebra de clases.

Boole desarrolla sus propuestas en su primer libro, *The Mathematical Analysis of Logic, being an essay towards a calculus of deductive reasoning* (1847) en el que propone un análisis de los enunciados tradicionales, A, E, I, O, en términos de ecuaciones y donde demuestra la validez de un silogismo mediante manejos algebraicos: si de las premisas, mediante manejos algebraicos, se puede obtener la conclusión, entonces, el silogismo es válido. El manejo que se da de los términos de los enunciados es mediante una interpretación como términos de clases; la siguiente tabla muestra la interpretación de Boole:

A:	Todo X es Y	$x(1-y)=0$
E:	Ningún X es Y	$xy=0$
I:	Algún X es Y	$xy\neq 0$
O:	Algún X no es Y	$x(1-y)\neq 0$

en donde la expresión '$(1-\alpha)$' representa el complemento de la clase 'α'. Así, la primera expresión se puede interpretar como que es vacía la intersección de la clase de las X y el complemento de la clase de las Y; la segunda expresión señala que es vacía la intersección de la clase de las X y la clase de las Y; la tercera expresión señala que *no es vacía* la intersección de la clase de las X y la clase de las Y y, finalmente, la cuarta expresión señala que *no es vacía* la intersección de la clase de las X y el complemento de la clase de las Y.

Los siguientes ejemplos de silogismos válidos muestran, de manera intuitiva, cómo se podrían emplear las ecuaciones (y las desigualdades), junto con razonamiento algebraico, para obtener las conclusiones deseadas:

Primera figura (bArbArA):

Todo C es D $\quad c(1-d)=O \therefore c=O \vee 1-d=O$

Todo B es C $\quad b(1-c)=O \therefore b=O \vee 1-c=O$

pero o bien $c=O$ ó $1-d=O$:

si $c=O$, entonces $1-c\neq O$, $\therefore b=O \vee b(1-d)=O$

si $1-d=O$, entonces $b(1-d)=O$

$\therefore$ en ambos casos, $b(1-d)=O$

esto es: Todo B es D

Segunda figura (bArOcO):

Todo B es C $\quad b(1-c)=O \therefore b=O \vee 1-c=O$

Algún D no es C $\quad d(1-c) \neq O \therefore 1-c\neq O,\ d\neq O$ y $b=O$

$\therefore$ Algún D no es B $\quad$ pues $d(1-b)\neq O$, ya que $d(1-b)=d(1-O)=d\neq O$

Primera figura (dArII):

Todo C es B $\quad c(1-b)=O \therefore c=O \vee 1-b=O$

Algún D es C $\quad dc\neq O \therefore c\neq O,\ d\neq O$ y $1-b=O \therefore b=1\neq O$

$\therefore$ Algún D es B $\quad$ pues $db=d\neq O$

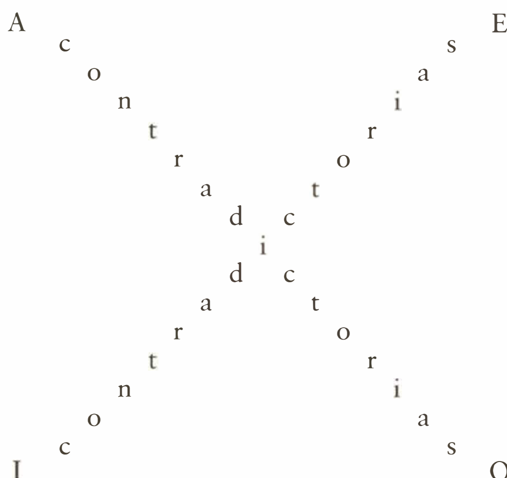
Las líneas anteriores hacen claro que mediante los manejos algebraicos de Boole no sólo se puede validar un argumento de forma silogística, sino que también se puede obtener una conclusión válida si sólo tenemos a la mano las dos premisas del silogismo.

Sin embargo, no es posible validar el silogismo aristotélico de la tercera figura (dArAptI), pues

	Sólo si $c \neq O$, entonces:
Todo C es B	$c(1-b)=O \therefore c=O \vee 1-b=O \;/\; b=1$
Todo C es D	$c(1-d)=O \therefore c=O \vee 1-d=O \;/\; d=1$
∴ Algún D es B	$db \neq O \therefore bd = 1 \times 1 \neq O$

Aquí nos enfrentamos a una interpretación de los enunciados universales que no les confiere *contenido o carga existencial.* Expresar, como lo hace Boole, que un enunciado universal afirmativo (A), 'Todo B es C', es algebraicamente representable como $b(1-c)=0$ o que uno universal negativo (E), 'Ningún B es C', es representable como $bc=0$, es expresar que la intersección de dos clases es vacía o que *no existen individuos* que tengan, conjuntamente, las propiedades *b* y $1-c$ o bien *b* y *c,* respectivamente. Pero, de esto *no* se puede inferir que *hay* individuos que sean *c* o miembros de la clase *c.*

Conforme a esta interpretación, que es la que se adopta en la lógica contemporánea, el que hemos presentado como cuadro de oposición aristotélico (*cf. supra,* p. 50), pierde las aristas laterales, esto es, si los enunciados universales, A y E, *no* tienen contenido existencial, pero sí lo tienen los enunciados particulares, I, 0 pues, en la versión algebraica de Boole, la representación de estos enunciados es mediante una desigualdad que señala *que no es vacía* la intersección de dos clases, esto es, (I) 'Algunos B son C', que se representa como $bc \neq 0$ y (O) 'Algunos B no son C', que se representa como $b(1-c) \neq 0$, entonces de la verdad de los enunciados universales *no* se sigue la verdad de los particulares correspondientes, esto es, desaparece la relación de subalternación. Pero, además, si la clase b es vacía, entonces es verdadero tanto que $b(1-c)=0$ como que $bc=0$, y será falso tanto que $b(1-0) \neq 0$ como que $bc \neq 0$. De esto se sigue que, a diferencia del cuadro de oposición aristotélico, ahora podrán ser *verdaderos a la vez* los enunciados universales, A y E, y *falsos a la vez* los enunciados particulares, I y O, por lo que ya no habrá más enunciados contrarios ni enunciados subcontrarios, así como tampoco valdrá la relación de subalternación. El cuadro, así, se ve reducido a sus dos grandes diagonales, esto es, sigue habiendo enunciados contradictorios: A y O, por una parte, y E e I por la otra:



En la segunda mitad del siglo XIX hubo una gran proliferación de nombres importantes para el desarrollo de la lógica. Por razones obvias de espacio, no podemos detenernos a considerar mínimamente a estos lógicos. Un pensador, sin embargo, debe mencionarse pues con él surge la lógica en su versión contemporánea: Gottlob Frege. Aquí señalaremos algunas de las aportaciones centrales de este hombre que creó la lógica de nuestros días.

IV. LA LÓGICA DE FREGE

Gottlob Frege (1848-1925), desarrolla un primer sistema axiomático, plenamente simbolizado, consistente y *completo,* de lógica de 1.er orden, aún antes de que se tuvieran las herramientas lógicas adecuadas para llevar a cabo la prueba de la completud de un sistema deductivo cualquiera. Como es bien sabido, el trabajo de Frege quedó, por un tiempo, fuera del cauce principal del desarrollo de la lógica debido, principalmente, a su muy rígico y estorboso sistema de notación. Por otra parte, el interés que tuvo Russell por su obra, con muchos puntos de contacto con la suya propia, y la difusión que de ella hizo, la pusieron en el primer plano de la atención filosófico-matemática de la Europa de los primeros años de este siglo y tal atención aún sigue fija en su labor ahora a casi setenta años de su muerte.

El interés que se tiene por la obra de Frege no sólo se refiere a su trabajo técnico matemático, sino también, y muy especialmente, a sus formulaciones filosófico-matemáticas acerca de diversos problemas tanto epistémicos como ónticos que rebasan el terreno relacionado con solamente los fundamentos de las matemáticas. Aquí vale la pena mencionar

su importante discusión de la distinción semántica entre sentido y referencia que, a partir de la atención que le prestaron Carnap y Church, ha sido estudiada con cuidado por un sinnúmero de filósofos posteriores.

El Begriffsschrift *y el origen de la lógica contemporánea*

En 1879, Frege publica una breve obra, la primera que dedica al campo de la lógica, su *Begriffsschrift* (1879), que se convertirá en la obra que marca el comienzo de la lógica formal contemporánea. En ella, como ya lo señalé, Frege formula un sistema de lógica de primer orden en el que su autor introduce una modificación radical en el análisis de las proposiciones, ya que, en lugar de analizarlas como si fueran de la forma sujeto-predicado, propone verlas bajo la forma de función y argumento y, además, en su escrito las pruebas se llevan a cabo de una manera estrictamente formal. Hay que añadir que el trabajo de Frege también se caracteriza por el rigor en la presentación de sus demostraciones, que no figura en obras posteriores como los *Principia Mathematica* de Russell y Whitehead, por ejemplo.

La preocupación de Frege y el propósito de su trabajo se encuentran claramente expresados en el 'Prefacio' del *Begriffsschrift.* Al considerar cuál sea la forma mejor de establecer la verdad de una proposición, nos dice lo siguiente:

> Obviamente, la manera más confiable de llevar a cabo una prueba es seguir la lógica pura; ésta es una forma que, al dejar de lado las características particulares de los sujetos, depende tan sólo de las leyes en las que se funda todo conocimiento. Conforme a esto, nosotros dividimos en dos tipos todas las verdades que requieren de una justificación, a saber, aquellas para las que la prueba puede llevarse a cabo de manera puramente lógica y aquellas que deben apoyarse en hechos de la experiencia.

Más adelante nos sigue diciendo:

> Para impedir que cualquier cosa intuitiva penetrase aquí desapercibida, tuve que poner todo mi esfuerzo en mantener la cadena de inferencias libre de huecos...

y, para superar los obstáculos que le imponía el lenguaje natural, nos sigue diciendo que eso lo

> ... llevó a la idea de la presente ideografía [*Begriffsschrift*]. Su primer propósito es, pues, proporcionarnos la prueba más confiable de la validez de una cadena de inferencias y señalar toda presuposición que intente colarse desapercibida, de tal manera que se pueda investigar su origen.

Otra virtud que Frege encuentra en su lenguaje simbólico es que, para los propósitos científicos para los que fue creado, el mismo facilitará el proceso de análisis y, si esto es así, de ello se seguirá una mayor facilidad

para descubrir nuevas verdades, esto es, propiciará un mayor avance de la ciencia.

Frege mismo señala la relación que su nuevo lenguaje puede tener con el lenguaje universal —la *characteristica universalis*— de Leibniz, aun cuando considera que el entusiasmo de éste fue demasiado y que, por esto, subestimó las dificultades a las que habría de enfrentarse la tarea de crear un lenguaje así.

Si, ciertamente, la búsqueda de un lenguaje simple, que facilite las pruebas lo acerca a Leibniz, la idea general del método tiene claras reminiscencias cartesianas de las *Reglas para la conducción del espíritu.*

Por otra parte, además de lo que anteriormente he señalado con respecto al sistema de Frege, es preciso recordar que, en el mismo, se puede expresar, de manera clara, la cuantificación múltiple.

El sistema de Frege contiene, como sus conectivas básicas, la negación y el condicional, definido a la manera de Filón, y el cuantificador que usa como primitivo es el universal.

Presento algunas expresiones formales en el simbolismo de Frege junto con su traducción al simbolismo de Peano:

A
B

expresa, según lo señala Frege, el juicio de que no sucede que A se niegue y B se afirme (esto es, que no es el caso que B sea verdadera y A falsa).

En la notación de Peano, la expresión anterior de Frege se convierte en el condicional $(B \supset A)$.

Frege	Peano
A B	$(B \supset -A)$
a A (a) e B (a, e)	$(x)\ ((y)\ B(x,y) \supset Ax)$
a A (a) e B (a, e)	$(x)\ ((\exists y)\ -B(x,y) \supset Ax)$

En el sistema de Frege (usando la notación de Peano) figuran, junto con la regla de derivación *modus ponens:*

$$\begin{array}{l} \varphi \supset \psi \\ \varphi \\ \hline \therefore \psi \end{array}$$

los siguientes nueve axiomas que cubren tanto los cálculos proposicional y de predicados, así como la teoría de la identidad:

1. $(\varphi \supset (\psi \supset \varphi))$ 2. $((\chi \supset (\psi \supset \varphi)) \supset ((\chi \supset \psi) \supset (\chi \supset \varphi)))$
8. $((\chi \supset (\psi \supset \varphi)) \supset (\psi \supset (\chi \supset \varphi)))$

28. $((\psi \supset \varphi) \supset (-\varphi \supset -\psi))$ 31. $(\varphi \supset --\varphi))$ 41. $(--\varphi \supset \varphi)$
52. $((a=b) \supset (f(a) \supset f(b)))$ 54. $(a=a)$ 58. $(x)f(x) \supset f(y)$

La numeración que aquí aparece es la que usa Frege en su escrito.

Frege, con su *Begriffsschrift* unifica lo que autores anteriores, a partir de Aristóteles habían propuesto por separado, la lógica de enunciados y la lógica de términos; por otra parte, introduce una teoría general de la cuantificación que resuelve muchos problemas a los que se habían enfrentado los lógicos medievales y, junto con las otras aportaciones señaladas con anterioridad, da nacimiento a la lógica contemporánea. Después de él, se intensifica la investigación en la teoría lógica y se diversifican los sistemas lógicos que toman como punto de partida la lógica clásica bivaluada, que Frege genera con su trabajo.

Aquí tan sólo apunto el bien conocido interés de Frege por fundar la matemática en la lógica, aspiración que con él comparte Russell. Lamentablemente, en 1903, Frege publica el volumen II de sus *Grundgesetze der Arithmetik* al que añade un *Postscriptum* en el que anuncia la paradoja, descubierta por Russell, que surge de sus sistemas, a saber, la paradoja de las clases que no son miembros de sí mismas, lo que le produce un profundo pesar, ya que la misma muestra que su trabajo no se puede proponer como una fundamentación adecuada de la matemática. No es posible, sin embargo, que aquí ampliemos estas breves observaciones, ya que las mismas rebasan propiamente el campo de la lógica elemental que es el tema que nos ocupa.

Finalmente, menciono que Frege mismo y otros pensadores, entre ellos Boole y Peirce, proponen la idea de una matriz de evaluación para los enunciados del cálculo proposicional y, más adelante, la idea la elaboran con mayor precisión Lukasiewicz, Post y Wittgenstein.

V. DESPUÉS DE FREGE

1. *Russell, Whitehead y los* Principia Mathematica

La obra monumental de Russell y Whitehead, *Principia Mathematica,* cuyos tres volúmenes se publicaron, respectivamente, en los años 1910, 1912 y 1913, puede verse como la conclusión de una de las propuestas centrales de Frege en su labor en fundamentos de la matemática: mostrar que la matemática puede fundarse en la lógica. Russell y Whitehead intentan evitar la paradoja en el sistema de Frege y llevar a cabo la tarea de mostrar que es posible derivar toda la matemática de la lógica. Russell, para enfrentarse a la paradoja mencionada, desarrolla su teoría de los tipos lógicos. Sin embargo, lo que Russell toma como el fundamento lógico, primeramente, va más allá de lo que es la lógica elemental o lógica de primer orden y, por otra parte, Kurt Gödel mostró que es imposible derivar toda la matemática de una base axiomática.

2. *Kurt Gödel*

En el año de 1930, un joven austriaco, Kurt Gödel (1906-1978), de 24 años presenta, como disertación doctoral (que luego publicará), la demostración, sobre un conjunto de axiomas de *lógica elemental,* de que a partir de este conjunto es posible derivar todas (completud) y sólo (corrección) las verdades lógicas. Esto parecía apoyar la propuesta de Hilbert de demostrar la consistencia de las matemáticas, pues de la corrección del sistema se sigue sin problemas su consistencia, pero el mismo Gödel, un año después, en 1931, muestra que una axiomatización lo suficientemente fuerte como para derivar de ella la aritmética elemental de los números naturales, *si* la misma es consistente, entonces será esencialmente incompletable, esto es, habrá verdades matemáticas que no será posible obtener como teoremas, esto es, lo que muestra Gödel es que no son equivalentes las nociones de verdad matemática y la de teorema o bien, que no hay una equivalencia entre los aspectos semántico y sintáctico de la matemática; además, Gödel demuestra que será imposible demostrar la consistencia de ese sistema.

3. *La lógica y los fundamentos de la matemática*

Para finalizar esta breve visión histórica de la lógica, vale la pena señalar la posición que la disciplina ocupa dentro de las diferentes doctrinas que se propusieron para dar cuenta de los fundamentos de la matemática.

Líneas atrás se señaló el interés de Frege, que Russell comparte, de mostrar que la matemática se funda directamente en la lógica o, de manera quizá más precisa, que no hay ninguna diferencia esencial entre lógica y matemática, ya que ésta es una continuación de la primera. Aquí es importante señalar la visión ontológica de esta posición, conocida con el nombre de *logicismo:* la matemática es un estudio descriptivo de una realidad de tipo platónico, por lo que los enunciados matemáticos deben de ser verdaderos de dicha realidad. Así pues, la matemática es una continuación de la lógica que se funda en un conjunto verdadero de axiomas.

Una propuesta diferente a la logicista es la *formalista,* enunciada por David Hilbert. Conforme a ella, se mantiene una visión similar a la logicista en tanto que no se considera que haya una diferencia esencial entre lógica y matemática, tan sólo que *no* se mantiene una posición reduccionista de la matemática con respecto a la lógica, sino que se propone que ambas se desarrollen conjuntamente a fin de mostrar que el sistema conjunto está libre de contradicción. Aquí se dejan de lado los aspectos semánticos de verdad de los axiomas y el criterio básico de corrección es uno *sintáctico,* esto es, que sea imposible derivar en el sistema tanto una fórmula, φ, como su contrapuesta sintáctica, $-\varphi$, que se interpretaría como la negación de φ.

Hilbert propuso un programa que procediera de manera gradual *(el programa de Hilbert)* para demostrar la consistencia de la matemática

pero, según lo señalé líneas atrás, su cumplimiento se ve frustrado con la demostración de Gödel de que esto es imposible.

Finalmente, la posición *intuicionista,* cuyo principal defensor fuera Luitzen Egbertus Jan Brouwer (1881-1967), considera que la lógica surge de un proceso de abstracción que se lleva a cabo en base a ciertas regularidades que se observan en el proceso mismo de desarrollar la matemática. Así, lógica y matemática son *dos* disciplinas que claramente se distinguen dentro de la perspectiva intuicionista. Aquí es importante hacer notar que la lógica intuicionista se desvía de la lógica clásica en tanto que aquélla *no* acepta, como una verdad lógica, el principio $\vdash (\varphi \supset --\varphi)$ de la lógica clásica.

4. *Otras lógicas*

Según lo señalé con respecto a la lógica intuicionista, ésta se separa de la lógica clásica al no aceptar todas las tesis que figuran en ésta. Por otra parte, también se distingue de la lógica clásica en tanto que acepta *tres* valores de verdad, en lugar de los dos únicos valores, característicos de la tradición clásica. De esta manera, tenemos un aumento en valores de verdad y esto es algo que Lukasiewicz elabora, a partir de 1917, esto es, una *lógica multivaluada* y, de esta manera, se abre la posibilidad de ampliar y generalizar el estudio de estas lógicas hasta llegar a sugerir el estudio de lógicas infinitamente valuadas.

En 1918, C. I. Lewis introduce una noción de implicación más fuerte que la material, la implicación estricta, relacionada con la implicación de Diodoro, según se señaló en su momento y, con ello, presenta un sistema de *lógica modal,* que amplía el repertorio de la lógica clásica.

Por el año de 1956 Gregorio Klimovski y, más adelante, Héctor Neri Castañeda, Carlos Alchourrón y Andrés Raggio comienzan, en América Latina, a hacer uso de la lógica de manera creativa, en el caso de la teoría de los conjuntos, el primero, en el caso de lógica deóntica, Castañeda y Alchourrón y en el estudio de funciones recursivas y pruebas constructivas el último.

En 1963, Newton C. A. da Costa crea un nuevo sistema de lógica, denominado actualmente *lógica paraconsistente,* que ha despertado el interés de muchos lógicos contemporáneos.

En estos avances de la lógica, post *Principia Mathematica,* se puede hacer la distinción señalada por Susan Haack entre lógicas que son rivales o las que son ampliaciones de la lógica clásica. El lector encontrará, en las otras selecciones de este volumen, material que le permitirá precisar y ampliar las breves notas que figuran en esta historia mínima de la lógica.

Antes de terminar y sabiendo de antemano que no es posible dar una enumeración mínimamente satisfactoria de los estudiosos de la lógica en nuestros países, me atrevo a mencionar a algunas de las figuras destacadas a las que no he aludido anteriormente.

Me referiré aquí a Vicente Ferreira da Silva como el primer autor de un libro de lógica moderna editado en Latinoamérica, *Elementos de lógica matemática* (São Paulo, 1940) y a Francisco Miró Quesada como el iniciador de los estudios de lógica simbólica en el Perú y en Iberoamérica, con la publicación de su *Lógica* (1946).

En México, Javier Sánchez Pozos ha realizado trabajos importantes en lógicas no clásicas, especialmente en lógicas relevantes; Adolfo García de la Sienra ha hecho aplicaciones de la lógica a teorías económicas y Raúl Orayen, además de dedicarse a trabajar en la enseñanza y la investigación en lógica ha publicado un importante libro de teoría lógica.

En España es posible mencionar nombres destacados del pasado inmediato y del presente que dedican su atención a esta disciplina; entre ellos están Alfredo Deaño (†), Jesús Mosterín, Manuel Sacristán, Lorenzo Peña, quien también realizó su trabajo en Iberoamérica, etc.

BIBLIOGRAFÍA

Adamson, R. (1911), *A Short History of Logic,* Dubuque, Edinburgh-London, 1963.

Alberto de Sajonia (1522), *Perutilis logica,* Venezia (v. e. A. Muñoz García, México, 1988).

Ashworth, E. J. (1974), *Language & Logic in the Post Medieval Period,* Kluwer Academic, Dordrecht.

Ashworth, E. J. (1978), *The Tradition of Medieval Logic and Speculative Grammar.*

Bayer, R. (1954), *Epistemologie et logique depuis Kant jusqu'a nos jours,* Paris.

Blakey, R. (1851), *Historical Sketch of Logic from the Earliest Time to the Present Day,* London.

Blanché, R. (1970), *La logique et son histoire d'Aristote à Russell,* A. Colin, Paris.

Blanché, R., «Histoire de la logique», en *Enciclopædia Universalis,* vol. 10, 49-52.

Bochenski, I. M. J. (1961), *A History of Formal Logic,* University of Notre Dame Press, Notre Dame (v. e. M. Bravo Lozano, *Historia de la lógica formal,* Madrid, Gredos, 1967).

Bochenski, I. M. J. (1968), *Ancient Formal Logic,* North Holland, Amsterdam-London.

Böhner, Ph. (1959), *Medieval Logic. An outline of its development from 1250-*c. *1400,* Manchester Univesity Press, Manchester.

Boll, M. y Reinhart, J. ([2]1965), *Les Étapes de la logique,* Paris, v. e. N. Sito, *Las etapas de la lógica,* Buenos Aires, 1961.

Bowne, G. D. (1963), *The Development of the Philosophy of Logic from 1880 to 1908,* Ann Arbor.

Brody, B. A. (1967), *The Rise of the Algebra of Logic,* Ann Arbor.

Cenáculo, M. do (1958), *Da História da Lógica,* Lisboa.

Dalla Chiara, M. L. (1983), *Logic in the 20th Century,* Scientia, Milano.

Díez y Lozano, B. ([2]1928), *Historia de la lógica,* Murcia.

Enriques, F. (1922), *Per la storia della logica,* Bologna (v. e. J. L. Angelis, *Para la historia de la lógica,* Buenos Aires, 1949).

Franck, A. (1838), *Esquisse d'une histoire de la logique,* Paris.

Geach, P. Th. (1968), *Reference and Generality, an Examination of Some Medieval and Modern Theories,* Ithaca.

Haack, S. (1974), *Deviant Logic,* Cambridge University Press, Cambridge.

Harms, Fr. (1881), *Geschichte der Logik,* Berlin.

Kneale, W. y Kneale, M. ([3]1966), *The Development of Logic,* Oxford (v. e. J. Muguerza, *El desarrollo de la lógica,* Tecnos, Madrid, 1980).
Kneebone, G. T. (1965), *Mathematical Logic and the Foundations of Mathematics. An Introductory Survey,* D. Van Nostrand & Co., New York, Toronto, Princeton.
Kotarbiński, T. (1964), *Leçons sur l'histoire de la logique,* Paris.
Largeault, J. (1970), *Logique et philosophie chez Frege,* Nauwelaerts, Paris, Louvaine, 1970.
Lewis, C. I. (1960), *A Survey of Symbolic Logic,* Dover, New York.
Lukasiewicz, J. ([2]1957), *Aristotle's Syllogistic, from the Standpoint of Modern Formal Logic,* Clarendon Press, Oxford.
Mackay, D. (1920), *Mind in the Parmenides. A Study in the History of Logic,* New York. Los Ángeles, 1924.
McCall, S. (1967), *Polish Logic 1920-1939,* Clarendon Press, Oxford.
Mercado, T. de (1571), *Comentarios lucidísimos al texto de Pedro Hispano,* introducción y traducción de M. Beuchot, Universidad Nacional Autónoma de México, México, 1986.
Miró Quesada, F. y Carrión Wam, R. (comps.) (1988), *Antología de la lógica en América Latina,* Universidad de Carabobo, Fundación Banco Exterior, Valencia, Venezuela-Madrid.
Mostowski, A. (1965), *Thirty Years of Foundational Studies. Lectures on the Development of Mathematical Logic and the Study of the Foundations of Mathematics in 1930-1964,* Helsinki.
Nidditch, P. H. (1962), *The Development of Mathematical Logic,* New York.
Patzig, G. (1968), *Aristotle's Theory of the Syllogism. A Logico-Philological Study of Book and of the Prior Analytics,* D. Reidel, Dordrecht-Boston.
Pedro Hispano, *Tractatus,* llamado después *Summule Logicales,* primera edición crítica basada en los manuscritos e introducción de L. M. de Rijk (v. e. M. Beuchot, México, 1986).
Prantl, C. (1855-1870), *Geschichte der Logik im Abendlande,* Leipzig-Darmstadt-Graz-Berlin, 1955.
Risse, W. (1964), *Die Logik der Neuzeit. 1. Band 1500-1640,* Stuttgart-Bad Cannstatt.
Robert, S. (1978), *La logique, son histoire, ses fondements,* Le Préambule, Longueuil.
Robles, J. A. (1980), «La generalización múltiple y la cuantificación en la lógica de Frege»: *Episteme, 4,* 37-42.
Scholz, H. (1931), *Geschichte der Logik,* Junder u. Dünnhaut, Berlin.
Scholz, H. (1968), *Esquisse d'une histoire de la logique,* Aubier-Montaigne, Paris.
Shearmann, A. Th. (1906), *The Development of Symbolic Logic,* London.
Van Evra, J. W. (1966), *A History of some Aspects of the Theory of Logic, 1850-present,* Ann Arbor.
Van Heijenoort, J. (comp.) ([3]1976), *From Frege to Gödel. A Source Book in Mathematical Logic, 1879-1931,* Harvard University Press, Cambridge.
Vuillemin, J. (1968), *Leçons sur la première philosophie de Russell,* A. Colin, Paris.
Ziehen, Th. (1920), *Lehrbuch der Logik auf positivistischer Grundlage mit Berücksichtigung der Geschichte der Logik,* Bonn.

LÓGICA CLÁSICA DE PRIMER ORDEN

Daniel Quesada

La *lógica clásica de primer orden* es la teoría más versátil y aplicable, también la más estudiada y mejor conocida, de la lógica contemporánea, y ocupa un lugar central en esta ciencia, siendo el «punto de referencia» para otras partes o teorías de la lógica. Otros nombres con los que se la conoce son *lógica de predicados (elemental)* y *lógica cuantificacional.* Puesto que esta monografía sólo trata de lógica clásica, en adelante prescindiremos en general de este adjetivo.

El objetivo más patente de una teoría lógica es ofrecer una explicación de la relación de *implicación lógica* en que se encuentran las premisas y la conclusión de una inferencia correcta. Otro objetivo es ofrecer un método sistemático para separar las inferencias correctas de las que no lo son. Al perseguir estos objetivos la lógica contemporánea ha concebido las inferencias como formuladas lingüísticamente y se ha servido, del modo que indicaremos, de *lenguajes artificiales* desarrollados precisamente para alcanzarlos. De entre éstos, precisamente la familia más importante es la de los *lenguajes de primer orden.*

La lógica (clásica) de primer orden abarca también en cierto sentido la *lógica (clásica) de enunciados* (o *lógica proposicional*). Los lenguajes utilizados en esta rama de la lógica muestran parte de los recursos lingüísticos de los lenguajes de primer orden. Por ello, a las exposiciones de la lógica de primer orden en sentido propio suele anteceder la de la lógica de enunciados, y también esta monografía comenzará por una exposición de esta rama básica de la lógica. Muchos de los conceptos presentados en este marco más restringido nos servirán luego en el más amplio.

I. FUNCIONES VERITATIVAS

La idea *intuitiva,* preteórica, que tenemos de una conexión lógica o relación de *implicación lógica* entre las premisas y la conclusión de un argu-

mento o una inferencia es la siguiente: es imposible que las premisas sean verdaderas y la conclusión sea falsa. Si ésta conexión se da, decimos que las premisas implican lógicamente la conclusión (o una variante terminológica de ello: que la conclusión se sigue lógicamente de las premisas; que es una *consecuencia lógica* de ellas, etc.).

No importan pues a la lógica las cualidades retóricas o estilísticas, o la capacidad persuasiva de los argumentos. Ni tampoco los procesos psicológicos o neurofisiológicos implicados en su producción. Todo lo que importa es la conexión, por el momento intuitivamente descrita, entre premisas (los enunciados que se dan por supuestos en el razonamiento) y conclusión.

Lo que necesitamos es tener una buena explicación de en qué consistiría o a qué se debería la mencionada imposibilidad de que en una inferencia correcta las premisas sean verdaderas y la conclusión falsa, y también, si ello es posible, un método que nos sirva para determinar si se da o no en un caso cualquiera.

Algunos fragmentos del discurso, o usos de oraciones en el lenguaje natural son *veritativo-funcionales.* Ello quiere decir que el *valor veritativo* (verdadero o falso) del fragmento en cuestión o de la oración compleja utilizada es una función de los valores veritativos de las oraciones simples que componen el discurso u oración compleja, es decir, depende de esos valores de una manera totalmente definida. Esta dependencia es, como veremos en la sección siguiente, de gran relevancia para el estudio de las propiedades lógicas fundamentales.

Los lenguajes de la lógica de enunciados están diseñados para estudiar sistemáticamente tales dependencias. Un lenguaje de éstos contiene los siguientes elementos:

1) Letras de enunciado: p, q, r, s, ...
2) Conectivas: $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$.
3) Paréntesis: (,).

El número de *letras de enunciado* de los lenguajes de la lógica de enunciados puede ser «arbitrariamente grande»: podemos considerar lenguajes con un número infinito numerable de ellas (tantas como números naturales: p_1, p_2, p_3, ...)[1]. Estas letras pueden utilizarse para representar oraciones simples (no compuestas veritativo-funcionalmente) o que, a los propósitos inmediatos, queramos considerar como tales.

Enseguida prestaremos una atención especial a las *conectivas.* Sus denominaciones más comunes son, respectivamente, *negación, conyunción, disyunción, condicional* y *bicondicional,* y, como estos nombres indican, los respectivos análogos de las expresiones «no», «y», «o», «si... entonces», «si y sólo si»[2]. Más adelante mencionaremos también la posibilidad de escoger otros conjuntos de conectivas.

1. En la actualidad se consideran incluso «lenguajes» con una cantidad no numerable de símbolos, pero esta posibilidad no será tenida en cuenta en la presente monografía, ni en este caso ni en el de primer orden.

2. Se utilizan a veces símbolos distintos para alguna de las conectivas. Así, puede encontrarse $\sim$ en lugar de $\neg$, & en lugar de $\wedge$, y $\supset$ en lugar de $\rightarrow$.

La función de los paréntesis es auxiliar: desambiguar sintácticamente. Otras convenciones pueden adoptarse al mismo fin.

A partir de los mencionados elementos se construyen *fórmulas* (bien formadas); por ejemplo: $\neg q$, $r \land p$, $p \rightarrow (\neg q \lor s)$. Recurriendo a sus análogos, se leen así, respectivamente: «no *q*», «*r* y *p*», «si *p,* entonces no-*q* o *s*». En cambio, otras cadenas de símbolos que, en principio, podríamos formar no constituyen fórmulas: $p \land, \neg 5(q \lor s)$ (como será obvio si se leen utilizando las expresiones análogas y teniendo en cuenta que las fórmulas pueden representar oraciones).

Las fórmulas de estos lenguajes se construyen sintácticamente de un modo preciso que excluye la ambigüedad (al contrario de lo que sucede con sus análogos en el lenguaje natural). Esto se realiza mediante una definición que delimita la clase de las fórmulas (frente a otras cadenas de símbolos no bien formadas).

Para dar tal definición nos servimos de otra serie de símbolos —aquí las letras griegas minúsculas Φ, Ψ, γ, con subíndices si es necesario— para representar, en principio, filas de símbolos cualesquiera. Debe tenerse en cuenta que estas letras no son símbolos de un lenguaje de la lógica de enunciados, sino símbolos de un lenguaje (concretamente: el español ampliado con tales símbolos) que utilizamos para caracterizar tal lenguaje. De forma totalmente general, en la terminología técnica, al lenguaje que en cada caso se utilice para describir o caracterizar otro lenguaje se le denomina *metalenguaje* (que es una noción relativa, puesto que en muchos casos pueden «cambiarse las tornas»: es posible tanto utilizar el español para describir el inglés como a la inversa).

Por lo dicho, las letras Φ, Ψ, γ, mientras no se especifique más, podrían representar una fila de signos cualquiera, por ejemplo, $\neg \lor rq \land$, o bien $r \land p$, el segundo de nuestros ejemplos anteriores de fórmulas, o una letra de enunciado. Mezclamos tales letras con símbolos de lenguaje de un modo fácil de entender (por simplicidad no utilizaremos otros recursos técnicos existentes para evitar toda posible confusión).

Con tales recursos, he aquí la definición que delimita la clase de las fórmulas:

1) Toda letra de enunciado es una fórmula.
2) Si Φ es una fórmula, también lo es $\neg \Phi$.
3) Si Φ, Ψ son fórmulas, también lo son $(\Phi \land \Psi)$, $(\Phi \lor \Psi)$, $(\Phi \rightarrow \Psi)$, $(\Phi \leftrightarrow \Psi)$.
4) Nada es una fórmula a menos que resulte de aplicar las cláusulas 1-3.

Ahora puede verse que $(p \rightarrow (\neg q \lor s))$, por ejemplo, es una fórmula. En efecto: *p, q, s* son fórmulas por la cláusula 1; $\neg q$ lo es entonces por la 2; por tanto, $(\neg q \lor s)$ lo es por la 3 y, finalmente, $(p \rightarrow (\neg q \lor s))$ es una fórmula, por la cláusula 3. Cada paso descansa en el resultado del anterior. Estas cláusulas determinan además un análisis sintáctico unívoco para cada fórmula y hacen que se pueda hablar del tipo sintáctico

de fórmula sin ambigüedad. Así, por ejemplo, una fórmula en cuya construcción la última cláusula aplicada es la del condicional, es un condicional. (Incidentalmente: la parte de una fórmula condicional a la izquierda del signo → se denomina *antecedente,* y la parte de la derecha, *consecuente*).

Normalmente se adoptan algunas convenciones para evitar el engorro que suponen los paréntesis cuando no son necesarios para realizar su tarea desambiguadora. Así, por ejemplo, se establece la convención de no escribir los paréntesis más externos al acabar de construir una fórmula (en nuestros primeros ejemplos de fórmulas se hizo ya uso de esa convención).

Este tipo de definición en el que unas cláusulas remiten a otras sin circularidad se denomina *definición recursiva.* Se le denomina «definición» porque, haciendo cierto uso de recursos técnicos, se la puede transformar en una definición explícita.

En adelante, utilizaremos Φ, Ψ, γ específicamente para fórmulas. Mostremos ahora la dependencia funcional de los valores veritativos. Esta se da en último término porque las conectivas expresan funciones de verdad, como revela la tabla (1.1).

Tabla (1.1)

Φ	¬ Φ
V	F
F	V

Φ	Ψ	(Φ∧Ψ)	(Φ∨Ψ)	(Φ→Ψ)	(Φ↔Ψ)
V	V	V	V	V	V
V	F	F	V	F	F
F	V	F	V	V	F
F	F	F	F	V	V

En (1.1) se presentan las *tablas de verdad* correspondientes a cada una de las conectivas de un lenguaje típico de la lógica de enunciados. Estas tablas muestran el modo en que tales conectivas contribuyen a determinar el valor veritativo (verdadero o falso, abreviadamente: V o F) de cualquier fórmula construida mediante las mismas.

Atendiendo a las posibilidades combinatorias que las tablas hacen patentes, vemos que hay más conectivas posibles. Concretamente, hay 16 conectivas binarias, pero bastan las que se han dado para desempeñar la función de las restantes, pues éstas pueden definirse en función de aquéllas. Incluso puede tomarse un conjunto más reducido de ellas. Puede demostrarse que, por ejemplo, basta tomar ¬ y ∨ como básicas, o ¬ y →.

Naturalmente, el valor veritativo de una fórmula no lo determinan *totalmente* las conectivas que en ella intervienen. Tal valor depende, además, de los valores veritativos que se asignen a sus letras de enunciado. La contribución de ambos factores puede apreciarse en la tabla de verdad de, por ejemplo, la fórmula $p \rightarrow (\neg q \vee s)$ (cf. 1.2).

Tabla (1.2)

p	q	s	$p\rightarrow($	$\neg q$	$\vee s)$
V	V	V	V	F	V
V	V	F	F	F	F
V	F	V	V	V	V
V	F	F	V	V	V
F	V	V	V	F	V
F	V	F	V	F	F
F	F	V	V	V	V
F	F	F	V	V	V
			3	1	2

En las tres columnas de la izquierda tenemos todas las maneras en que es posible asignar valores veritativos a las letras de enunciado de la fórmula. En cada una de las filas correspondientes a esa parte izquierda tenemos una asignación distinta. La asignación se toma en el sentido matemático de *función:* asignamos a *cada* cosa de un tipo exactamente *una* cosa de otro (aquí, a cada letra de enunciado, exactamente un valor veritativo). De manera que tenemos, para tres letras de enunciado, un total de ocho de tales asignaciones (en general, para n letras, 2^n asignaciones).

En las tres columnas de la derecha se reflejan los valores veritativos que adquieren las diversas subfórmulas de la fórmula de nuestro ejemplo y también ésta misma. Los números (que se han puesto aquí para ayudar a la explicación, pero que son inesenciales) indican el orden en el proceso. En primer lugar se calculan los valores veritativos de la columna 1 teniendo en cuenta los valores de la fórmula q (dados en una de las columnas de la izquierda, pues es una letra de enunciado) y la contribución de la conectiva $\neg$ (dada en la tabla de verdad para esta conectiva en (1.1)). Fijémonos por ejemplo, en la tercera fila. En ella el valor de q es F, por lo que el valor de $\neg q$, obtenido consultando la tabla (1.1), debe ser V, y éste es, en efecto, el valor que aparece para esta fórmula en la tercera fila de la columna correspondiente a $\neg q$. Los valores así calculados para $\neg q$ se utilizan ahora para, junto con los de la fórmula s obtener los de la fórmula $(\neg q \vee s)$ (columna número 2), de acuerdo con la tabla para la conectiva $\vee$ (ver tabla (1.1)). Finalmente, los valores de $p\rightarrow(\neg q \vee s)$ se obtienen (columna 3) de los de p y de los ya obtenidos para $(\neg q \vee s)$, de acuerdo con la tabla para $\rightarrow$.

En la sección siguiente se amplía la consideración de estas nociones y se examina su relevancia para la explicación de la «conexión lógica» entre las premisas y la conclusión de una inferencia.

II. LÓGICA DE ENUNCIADOS Y PROPIEDADES LÓGICAS

Sea L un lenguaje de la lógica de enunciados. Además de las conectivas (y de los paréntesis, símbolos meramente auxiliares), L tiene letras de

enunciado. Las fórmulas de *L* se obtienen por medio de una definición como la del apartado anterior. Sea $\mathbf{A}$ una asignación de los valores veritativos V o F a *todas las letras de L.* A partir de $\mathbf{A}$ se obtiene unívocamente una asignación de valores veritativos a *todas las fórmulas de L.* Ello se desprende de la *definición de verdad* para *L.*

Para presentar esta definición es conveniente introducir algo de notación. Cuando Φ sea una letra de enunciado, abreviaremos «$\mathbf{A}$ asigna a Φ el valor V» mediante $\mathbf{A}(\Phi) = V$. La definición toma la forma de una definición recursiva de «Φ es *verdadera en* $\mathbf{A}$», que abreviamos así: $\| \Phi \|_{\mathbf{A}} = V$, expresión en la que Φ puede ser una fórmula cualquiera (letra de enunciado o fórmula compleja)[3]. Como también es usual, abreviamos «Φ no es verdadera en $\mathbf{A}$» mediante $\| \Phi \|_{\mathbf{A}} \neq V$.

(2.1) Para cualquier letra de enunciado Φ de *L,* $\| \Phi \|_{\mathbf{A}} = V$ si y sólo si $\mathbf{A}(\Phi) = V$.

Para cualesquiera fórmulas Φ, Ψ de *L,*
$\| \neg \Phi \|_{\mathbf{A}} = V$ si y sólo si $\| \Phi \|_{\mathbf{A}} \neq V$;
$\| (\Phi \wedge \Psi) \|_{\mathbf{A}} = V$ si y sólo si $\| \Phi \|_{\mathbf{A}} = V$ y $\| \Psi \|_{\mathbf{A}} = V$;
$\| (\Phi \vee \Psi) \|_{\mathbf{A}} = V$ si y sólo si $\| \Phi \|_{\mathbf{A}} = V$ o $\| \Psi \|_{\mathbf{A}} = V$ o las dos cosas;
$\| (\Phi \rightarrow \Psi) \|_{\mathbf{A}} = V$ si y sólo si $\| \Phi \|_{\mathbf{A}} \neq V$ o $\| \Psi \|_{\mathbf{A}} = V$;
$\| (\Phi \leftrightarrow \Psi) \|_{\mathbf{A}} = V$ si y sólo si, o bien $\| \Phi \|_{\mathbf{A}} = V$ y $\| \Psi \|_{\mathbf{A}} = V$, o bien $\| \Phi \|_{\mathbf{A}} \neq V$ y $\| \Psi \|_{\mathbf{A}} \neq V$.

Podemos definir ahora $\| \Phi \|_{\mathbf{A}} = F$ (Φ es falsa en la estructura $\mathbf{A}$) simplemente como: $\| \Phi \|_{\mathbf{A}} \neq V$.

Introduzcamos algo más de terminología. En el contexto de la lógica de enunciados, a una asignación de valores veritativos a las fórmulas de un lenguaje *L* le llamamos una *estructura para L.* Si una fórmula Φ es verdadera en una estructura $\mathbf{A}$ (es decir, si $\| \Phi \|_{\mathbf{A}} = V$), decimos que $\mathbf{A}$ es un *modelo de* Φ. Consideremos ahora un conjunto cualquiera de fórmulas de un lenguaje *L;* llamémosle Γ. Si todas las fórmulas de Γ son verdaderas en una estructura $\mathbf{A}$, decimos entonces que $\mathbf{A}$ es un modelo de Γ.

Las nociones de *estructura para un lenguaje* y de *modelo de una fórmula o conjunto de fórmulas* son nociones clave de la lógica actual. Aquí las encontramos en el contexto restringido de la lógica de enunciados, pero más adelante las definiremos en el contexto más amplio de la lógica de primer orden.

Introducimos el signo $\models$ para la noción precisa de implicación lógica. De manera que leemos

(2.2) $\Gamma \models \Phi$

como: «El conjunto de fórmulas Γ *implica lógicamente* la fórmula Φ»; o también: «Φ es *consecuencia lógica* de Γ». La definición de este concepto es la siguiente:

3. Una notación alternativa muy utilizada es $\mathbf{A} \models \Phi$. Esta notación tiene el inconveniente de hacerle desempeñar una doble función al símbolo $\models$ (véase algo más adelante la otra función de este símbolo).

(2.3) $\Gamma \models \Phi$ si y sólo si todo modelo de Γ es también un modelo de Φ.

Más adelante abordaremos la importante cuestión de cuál es la relación entre la *noción precisa* que acabamos de introducir y la *noción intuitiva* que mencionábamos en el apartado anterior. Ahora aclararemos el sentido de esta definición al hilo de un ejemplo.

¿Qué quiere decir que la relación $\models$ de implicación lógica se da entre el conjunto formado por las dos fórmulas $r \rightarrow (\neg q \rightarrow p)$ y $r \wedge \neg q$, y la fórmula p? Expresado con símbolos:

(2.4) $r \rightarrow (\neg q \rightarrow p),\ r \wedge \neg q \models p$

Aplicando la definición (2.3) y atendiendo también al significado otorgado anteriormente a la noción de *modelo,* lo que ello quiere decir es que si en una estructura cualquiera las primeras dos fórmulas son verdaderas, también lo es la fórmula p.

Ahora que comprendemos el sentido o significado que damos a la noción de *implicación lógica* resulta que también tenemos, en la lógica de enunciados, un *método* para averiguar si esa relación se da en un caso cualquiera. Esto se desprende de que una estructura es aquí también una asignación de valores veritativos a las letras de enunciado de un lenguaje, y bastará entonces considerar las estructuras que difieren en lo que asignan a las letras de enunciado de las fórmulas que estamos considerando. Es decir, aplicamos el siguiente

TEOREMA 2.1 *(Teorema de coincidencia):* Si **A** y **B** son estructuras que asignan los mismos valores veritativos a cada una de las letras de enunciados de una fórmula Φ, entonces $\mathbf{A} \models \Phi$ si y sólo si $\mathbf{B} \models \Phi$.

Según esto, en nuestro ejemplo bastará considerar estructuras que asignen valores distintos a las letras p, q, r. Las otras estructuras son, a los efectos pertinentes, equivalentes a una de tales estructuras. Pero justamente son las *tablas de verdad* las que recogen las asignaciones distintas. Construyamos, por tanto, una tabla de verdad (cf. 2.5) para las fórmulas del ejemplo anterior.

Tabla (2.5)

p	q	r	$r \rightarrow (\neg q \rightarrow p)$			$r \wedge \neg q$	
V	V	V	V	F	V	F	F
V	V	F	V	F	V	F	F
V	F	V	V	V	V	V	V
V	F	F	V	V	V	F	V
F	V	V	V	F	V	F	F
F	V	F	V	F	V	F	F
F	F	V	F	V	F	V	V
F	F	F	V	V	F	F	V

Tenemos ahora la situación siguiente. La parte izquierda de cada una de las filas de la tabla describe una de las estructuras relevantes. Por ejem-

plo, en la segunda fila se describe una estructura que asigna V a las letras de enunciado p y q, asigna F a r y asigna no importa qué a las restantes letras de enunciado que pudiera haber en el lenguaje. Consideremos ahora una fila cualquiera y una cualquiera de las fórmulas de la tabla, Φ; la tabla le da a Φ el valor V en esa fila si $\| \Phi \|_{\mathbf{A}} = V$, donde **A** es una cualquiera de las estructuras asociadas con esa fila. Si $\| \Phi \|_{\mathbf{A}} \neq V$, entonces la tabla le da a Φ el valor F. Por ejemplo, como puede comprobarse utilizando la definición (2.1), si **A** es una estructura que asigna V a p y q, y asigna F a r, entonces $\| r \rightarrow (\neg q \rightarrow p) \|_{\mathbf{A}} = V$, y, en efecto, si examinamos en la tabla el valor de esta fórmula en la segunda fila, comprobamos que éste es V. Tal es la estrecha relación que hay entre la *definición de verdad* (2.1) y las tablas de verdad.

Siguiendo con nuestro ejemplo, concentremos nuestra atención en la fila 3; ésta resulta ser la única fila en que las dos fórmulas $r \rightarrow (\neg q \rightarrow p)$ y $r \wedge \neg q$ son verdaderas simultáneamente. Pero p también es verdadera en esa fila. Sucede pues que en todas las filas en las que las dos primeras fórmulas son verdaderas también lo es p. Por la relación que se acaba de explicar entre filas de una tabla y estructuras, podemos ahora decir que en toda estructura en que las dos primeras fórmulas son verdaderas, p lo es también. O, dicho de otro modo, todo modelo de las dos primeras fórmulas es también un modelo de p. Es decir, por la definición (2.3), el conjunto de esas dos fórmulas implica lógicamente la fórmula p.

Es claro que este procedimiento lo podemos generalizar a cualquier caso en que se trata de averiguar si $\Gamma \models \Phi$, siempre que Γ sea un conjunto finito de fórmulas (de lo contrario la tabla no se acabaría nunca de construir). Puede haber limitaciones prácticas —si hay muchas fórmulas implicadas o éstas son muy complejas—, pero no hay límites de principio. En tal sentido las tablas de verdad nos suministran un *procedimiento de decisión.*

Existe también la posibilidad de que una fórmula Φ sea verdadera en toda estructura, lo que escribimos así: $\models \Phi$. Este caso puede considerarse un caso particular del caso general recogido en la definición (2.3), a saber el caso en que Γ es el conjunto vacío. Para este caso existe también una terminología especial; al menos en el contexto de la lógica de enunciados se dice entonces que Φ es una *tautología.* La noción de *verdad en toda estructura* es el concepto preciso que se corresponde con la noción intuitiva de *verdad lógica.*

Naturalmente, también pueden utilizarse las tablas de verdad para averiguar si una fórmula cualquiera dada es o no una tautología (se construye la tabla de verdad para esa fórmula y se examina si la misma tiene el valor V en *todas* las filas)[4].

4. Las tablas de verdad fueron introducidas por primera vez por C. S. Peirce en Peirce, 1902; también se utilizan en el *Tractatus* de Wittgenstein (1921), al parecer de manera independiente. «Tautología» es un término introducido por Wittgenstein, aunque en el *Tractatus* tiene un sentido vinculado a la *teoría de la figura* que es la teoría del significado que en tal obra se defiende.

He aquí una lista de algunas tautologías conocidas, a las que aún se suele aplicar el antiguo rótulo de *leyes lógicas:*

(2.6) *Lista de tautologías:*

— Las leyes conmutativa y asociativa para $\wedge$, $\vee$, $\leftrightarrow$; por ejemplo: $(\Phi\wedge\Psi)\leftrightarrow(\Psi\wedge\Phi)$.

— Algunas propiedades de la negación:

$\neg\neg\Phi\leftrightarrow\Phi$.

$\neg(\Phi\rightarrow\Psi)\leftrightarrow(\Phi\wedge\neg\Psi)$.

— Leyes de De Morgan:

$\neg(\Phi\wedge\Psi)\leftrightarrow(\neg\Phi\vee\neg\Psi)$.

$\neg(\Phi\vee\Psi)\leftrightarrow(\neg\Phi\wedge\neg\Psi)$.

— Otras leyes:

Tercio excluso: $\Phi\vee\neg\Phi$.

(No) contradicción: $\neg(\Phi\wedge\neg\Phi)$.

Contraposición: $(\Phi\rightarrow\Psi)\leftrightarrow(\neg\Psi\rightarrow\neg\Phi)$.

Exportación: $((\Phi\rightarrow\Psi)\rightarrow\gamma)\leftrightarrow((\Phi\wedge\Psi)\rightarrow\gamma)$.

Las tautologías de esta lista se dan en forma esquemática. Recuérdese que Φ, Ψ, γ, pueden ser cualquier fórmula, por lo que, en realidad, estos esquemas suministran infinitas tautologías.

El siguiente teorema relaciona los casos de implicación lógica con las tautologías:

TEOREMA 2.2 *(Teorema de la deducción):* $\Gamma, \Phi\vDash\Psi$ si y sólo si $\Gamma\vDash\Phi\rightarrow\Psi$.

Aplicando este teorema se puede, por ejemplo, obtener implicaciones lógicas a partir de las tautologías de la lista anterior. Tomemos la «ley» de exportación; ésta, como la mayoría de las de la lista, está formulada mediante un bicondicional ($\leftrightarrow$). Ahora bien, éste se «desdobla» en dos condicionales, obteniéndose así:

$\vDash(\Phi\rightarrow\Psi)\rightarrow(\neg\Psi\rightarrow\neg\Phi)$;

$\vDash(\neg\Psi\rightarrow\neg\Phi)\rightarrow(\Phi\rightarrow\Psi)$.

Aplicando el teorema 2.2 a la primera de estas dos tautologías, por ejemplo, se obtiene la implicación:

$\Phi\rightarrow\Psi\vDash(\neg\Psi\rightarrow\neg\Phi)$.

También se puede aplicar el teorema 2.2 para obtener, a la inversa, tautologías a partir de implicaciones. Por ejemplo, a partir de (2.4) se llega a una tautología, aplicando dos veces el teorema. En un primer paso se obtiene:

$r\rightarrow(\neg q\rightarrow p)\vDash(r\wedge\neg q)\rightarrow p$;

y, finalmente:

$\vDash(r\rightarrow(\neg q\rightarrow p))\rightarrow((r\wedge\neg q)\rightarrow p)$.

A esta tautología podría aplicársele ahora la regla o ley de exportación para obtener otra. Pero baste esto como ejemplo de las múltiples transformaciones que se hacen posibles aplicando el teorema 2.2.

Puede defenderse que la lógica de enunciados nos suministra una buena teoría lógica para los fragmentos veritativo-funcionales del len-

guaje, dándonos una explicación de las propiedades lógicas en ese contexto. Para verlo, comenzaremos por utilizar un ejemplo de inferencia o razonamiento sencillo como el siguiente:

(2.7) La inflación aumenta a menos que aumente la productividad si aumentan los salarios. Los salarios aumentan pero la productividad no. Por lo tanto, la inflación aumenta.

(El simplismo de este argumento económico no importa desde un punto de vista lógico, como no importaría el posible refinamiento de otro razonamiento que por su complejidad sería aquí tal vez menos adecuado como ejemplo.)

Representando las oraciones «la inflación aumenta», «la productividad aumenta», «los salarios aumentan» mediante letras de enunciado, por ejemplo, p, q y r, respectivamente, se puede «modelar» lo que de lógicamente relevante tiene el razonamiento, tomando como premisas las fórmulas $r \rightarrow (\neg q \rightarrow p)$ (para la oración hasta el primer punto) y $r \wedge \neg q$ (la oración entre el primer punto y el segundo), y como conclusión a p. Afirmar entonces que el razonamiento es correcto equivaldría entonces a hacer la precisa afirmación (2.4), afirmación que, como se ha visto anteriormente, es verdadera.

La lógica de enunciados tiene pues la capacidad de ser utilizada como teoría lógica para mostrar la corrección o la incorrección de un número indefinido de argumentos que responden a un número indefinidamente grande de esquemas de argumentación distintos, trascendiéndose en este sentido la labor de la lógica tradicional que se limitaba a presentar un pequeño número de esquemas de argumentación correctos[5].

Con todo, la capacidad de la lógica de enunciados para dar cuenta de la inferencia es muy limitada. Es más justo hacer su evaluación cuando se la considera parte de la lógica de primer orden, pero plantear la cuestión del primer modo ayuda a darse una idea, en un contexto más sencillo, de qué es lo que está implicado en el tema.

Como puede verse por el ejemplo anterior, la utilización de la lógica de enunciados como teoría lógica depende de dos pasos cruciales. En primer lugar, de la «modelización» en el sentido apuntado, es decir, la representación esquemática en el lenguaje artificial de las oraciones del lenguaje natural que componen la inferencia (también llamada *formalización*). En segundo lugar supone la opción por el concepto preciso de *implicación lógica* que se presentó anteriormente.

El primer paso depende, para empezar, de que podamos considerar que las oraciones del fragmento de lenguaje que nos interesa sean verdaderas o falsas, o que, al menos, sea una idealización razonable considerarlas así. Esta dependencia es clara, pues, como hemos visto, en la lógica clásica de enunciados se contemplan sólo esos dos valores veritativos como valores posibles para las fórmulas. Hay lógicos que señalan aquí una limi-

5. La superación de la lógica tradicional en este sentido concreto puede señalarse por vez primera en Boole (1847) y Boole (1854).

tación de la lógica clásica y han trabajado para ampliarla con la admisión de otro u otros valores veritativos (cf. la monografía sobre «Lógica multivalente» en este mismo volumen).

Pero ese primer paso depende además de que consideremos a expresiones como «no», «y», «pero», «o», «si...», «si... entonces», «a menos que» y otras expresiones similares del lenguaje natural como adecuadamente representadas por las conectivas. Esto ha sido vivamente discutido en algunos casos, especialmente en el caso de «si... entonces» y su representación mediante → (el llamado *condicional material*). Se ha argumentado que «si... entonces» supone a menudo una conexión «más fuerte» que la veritativo-funcional que se recoge mediante el condicional material. Pero también hay buenas réplicas, que explican las discrepancias intuitivas apelando a la distinción entre el contenido semántico estricto de una oración usada en un contexto y los principios pragmáticos que rigen la conversación[6].

Respecto al segundo paso, ¿qué es lo que justificaría la adopción del concepto preciso definido en (2.3)? No sólo, por supuesto, que sea un concepto preciso, que esté apoyado por un aparato matemático, pues no se trata de cambiar de tema; la noción precisa ha de tener un estrecho vínculo conceptual con la noción intuitiva de implicación lógica, caracterizada en los términos modales de posibilidad, imposibilidad o necesidad (cf. el comienzo de la sección I). Aquí la línea de justificación consistiría en defender, primero, que las inferencias correctas son *inferencias analíticas* (inferencias realizadas en virtud del significado de las oraciones componentes), o, tal vez, un subconjunto de ellas: inferencias que tienen en cuenta exclusivamente la *contribución semántica* (contribución al significado) de expresiones que juegan un papel especial en el discurso (para el fragmento del lenguaje que ahora estamos considerando, expresiones como «no», «y», «o», «si... entonces», etc.; y, segundo, que las conectivas de la lógica de enunciados recogen precisamente tal contribución, pues ésta estriba en la manera peculiar en que cada una de esas expresiones produce un valor veritativo para las oraciones compuestas con ella a partir de los valores veritativos de las oraciones componentes.

Algo análogo podría decirse de la relación entre la noción intuitiva de *verdad lógica* y la noción precisa de *verdad en toda estructura*. Pero el lugar para tratar estas cuestiones con mayor extensión y profundidad es el volumen de la Enciclopedia dedicado a la Filosofía de la lógica.

III. CÁLCULOS LÓGICOS

Un cálculo lógico es un sistema para obtener todas las tautologías e implicaciones lógicas, a partir de un conjunto de las primeras, que se seleccio-

6. Para las críticas de la lógica contemporánea respecto a estos puntos cf. Strawson, 1952. Para la mencionada defensa, Grice, 1989, capítulos 2, 3 y 4, y también Thomson, 1990. Véanse también las monografías sobre lógicas no clásicas del presente volumen.

nan como axiomas del cálculo, y/o una serie de reglas cuya aplicación sólo atiende a la *forma* de las fórmulas a las que se aplican. Durante un período importante del desarrollo de la lógica contemporánea, los cálculos lógicos tuvieron un protagonismo mayor que las consideraciones semánticas en que se ha basado la línea de exposición de las secciones anteriores, pero aquí prescindiremos casi totalmente de consideraciones históricas, remitiendo a la monografía sobre historia de la lógica de este mismo volumen.

La taxonomía más al uso sobre los cálculos los clasifica en *cálculos axiomáticos, cálculos de deducción natural* y *cálculos de secuentes*[7].

Los *cálculos axiomáticos,* también llamados cálculos «estilo Hilbert» o sistemas «estilo Frege-Hilbert», consisten en una serie de fórmulas a las que se llama precisamente *axiomas,* y unas reglas (que se reducen normalmente a una o dos en el caso de la lógica de enunciados), llamadas *reglas de derivación* o *deducción* (también *reglas de inferencia,* aunque esta denominación es menos adecuada). Como ejemplo, el siguiente cálculo se debe a Hilbert. Los axiomas son:

(A1) $\Phi \rightarrow (\Psi \rightarrow \Phi)$;

(A2) $(\Phi \rightarrow (\Psi \rightarrow \gamma)) \rightarrow ((\Phi \rightarrow \Psi) \rightarrow (\Phi \rightarrow \gamma))$;

(A3) $(\neg \Psi \rightarrow \neg \Phi) \rightarrow (\Phi \rightarrow \Psi)$.

La única regla del cálculo es la llamada *regla de separación* o, en terminología más tradicional, *modus ponens:*

$$\frac{\Phi \rightarrow \Psi \qquad \Phi}{\Psi}$$

Axiomas y reglas se dan en forma esquemática, de manera que admiten infinitas ejemplificaciones (cf. sección 2). El cálculo se presenta de modo que $\neg$ y $\rightarrow$ se toman como las únicas conectivas primitivas o no definidas.

Los axiomas (o sus ejemplificaciones) son todos ellos tautologías (por ejemplo: (A3) es una dirección de la «ley» de contraposición presentada antes) y las fórmulas obtenidas a partir de ellos por la aplicación de la regla, los llamados *teoremas,* también lo son. Sin embargo, aunque esto pueda jugar un papel en la motivación del cálculo (se pretende recoger en él todas las «verdades lógicas»), no juega ninguno en su presentación y funcionamiento. Para obtener teoremas cada uno de los pasos es o bien la ejemplificación de un axioma o bien se obtiene de las fórmulas obtenidas en pasos anteriores mediante la aplicación de la regla de derivación. Veamos, por ejemplo, la derivación o prueba formal de un teorema muy simple: $(\Phi \rightarrow \Phi)$:

1) $(\Phi \rightarrow ((\Phi \rightarrow \Phi) \rightarrow \Phi)) \rightarrow ((\Phi \rightarrow (\Phi \rightarrow \Phi)) \rightarrow (\Phi \rightarrow \Phi))$

7. Cf. Sundholm, 1983, que constituye una buena referencia para la presentación comparada de los diferentes tipos de cálculos. Otra excelente exposición, más breve y elemental, se encontrará en Hodges, 1983, secciones 6 y 7.

2) $(\Phi\rightarrow((\Phi\rightarrow\Phi)\rightarrow\Phi))$
3) $(\Phi\rightarrow(\Phi\rightarrow\Phi))\rightarrow(\Phi\rightarrow\Phi)$
4) $(\Phi\rightarrow(\Phi\rightarrow\Phi))$
5) $(\Phi\rightarrow\Phi)$.

La fórmula de la primera línea en esta prueba es una ejemplificación (a su vez esquemática) de (A2) en la que Φ y γ se han ejemplificado por Φ, y Ψ por $(\Phi\rightarrow\Phi)$. La fórmula de la 2 es una ejemplificación de (A1) ($\Phi\rightarrow\Phi$ por Ψ). La de la 3 se obtiene de las dos anteriores por aplicación de la regla. La fórmula de la línea 4 es nuevamente una ejemplificación de (A1). Y, por aplicación de la regla a las fórmulas de las líneas 3 y 4 se obtiene finalmente el teorema que queríamos probar.

Si se quiere utilizar un cálculo de este tipo para derivar una fórmula a partir de unos supuestos que actúen como premisas, entonces hemos de tratar a éstos como si fueran axiomas adicionales.

Como puede verse por la ilustración anterior, derivar fórmulas en un cálculo axiomático es algo sumamente poco intuitivo que rápidamente se transforma en una tarea muy compleja. Esta es la gran desventaja de este tipo de cálculos. Pero, si bien los cálculos axiomáticos no son adecuados para probar cosas *en* ellos, sí lo son, por su reducido «bagaje» inicial, para probar cosas *sobre* ellos. Además, tienen otra ventaja en el caso de las *lógicas no clásicas,* por la facilidad de «localización» (en los axiomas) de las diferencias entre unos y otros sistemas.

Durante varias décadas, los cálculos axiomáticos fueron los únicos existentes, de manera que el trabajo pionero de lógicos como Frege y Russell se desarrolló en su marco[8]. En el clásico Gentzen (1934) se introdujeron los llamados *cálculos de deducción natural.* Diseñados para efectuar pruebas formales de la manera más sencilla (más «natural») posible, estos cálculos se encuentran en múltiples variantes y están profusamente representados en los libros de texto[9].

Los cálculos de deducción natural incluyen en torno a una decena de reglas que, como la del *modus ponens,* tienen un carácter muy intuitivo. El modo general de proceder en la deducción recoge también maneras intuitivas de demostrar. Por ejemplo, si lo que se quiere es deducir una fórmula condicional, se puede adoptar el antecedente como supuesto auxiliar, siendo el nuevo objetivo derivar el consecuente. O se puede iniciar una deducción por *reducción al absurdo,* suponiendo la negación de la fórmula que se quiere derivar; si se llega entonces a una contradicción a partir de ese supuesto auxiliar, se considera deducida la fórmula en cuestión.

8. Cf. Frege, 1879-1893 y Whitehead y Russell, 1910, tres de los hitos más importantes en el desarrollo de la lógica contemporánea.

9. Algunas referencias: Anderson y Johnstone, 1962 (cf. Deaño, 1978); Garrido, 1981; Guttenplan, 1986; Kalish y Montague, 1980 (cf. Mosterín, 1983); Mates, 1987; Quine, 1981; Sacristán, 1973 y Suppes, 1975. En realidad, la mayoría de los autores presentan un cálculo de la lógica de enunciados integrado en uno de primer orden.

En el mismo artículo citado, Gentzen introdujo también los *cálculos de secuentes*. Como una breve explicación de este tipo de cálculos en su forma original resultaría poco informativa, centraremos nuestra atención en un «descendiente», los *cálculos analíticos,* exponiendo en detalle uno de ellos.

El cálculo que presentaremos constituye una variante de las llamadas *tablas semánticas*[10]. Como todos los cálculos analíticos, estos cálculos están muy estrechamente ligados a la siguiente consideración semántica: Si $\Gamma \models \Phi$, no hay ninguna estructura que haga simultáneamente verdaderas a las formulas de Γ y falsa a Φ. Pero esto quiere decir que no hay ninguna estructura que haga simultáneamente verdaderas a las fórmulas de Γ y a $\neg \Phi$, pues Φ es falsa en una estructura si y sólo si $\neg \Phi$ es verdadera en ella.

Un conjunto de fórmulas para el que hay al menos una estructura en la que todas las fórmulas del conjunto son verdaderas se llama *satisfacible,* en caso contrario, el conjunto es *insatisfacible.* Así, si $\Gamma \models \Phi$, el conjunto de fórmulas que resulta de añadir $\neg \Phi$ a las de Γ es insatisfacible.

Tomemos de nuevo un ejemplo. Supongamos que se trata de demostrar lo siguiente:

(3.1) $p \rightarrow (\neg q \wedge r),\ p \wedge \neg q \models r.$

Por lo que se acaba de decir, ello equivale a demostrar que el conjunto de las tres fórmulas siguientes es insatisfacible:

(3.2) $p \rightarrow (\neg q \wedge r),\ p \wedge \neg q, \neg r.$

Veamos si podemos *refutar* esa afirmación mostrando que ese conjunto es satisfacible, es decir, que hay al menos una estructura en que estas fórmulas son verdaderas. En primer lugar, para que $p \wedge \neg q$ sea verdadera, lo han de ser p y $\neg q$. Por otro lado, para que $p \rightarrow (\neg q \wedge r)$ sea verdadera, una al menos de estas dos cosas debe suceder: que p sea falsa (y, por lo tanto, $\neg p$ verdadera), o que $(\neg q \wedge r)$ sea verdadera. La primera de estas alternativas entra en conflicto con la exigencia anterior de que p sea verdadera. Pero para que la segunda alternativa se dé, tanto $\neg q$ como r deben ser verdaderas, y esto es imposible porque $\neg r$ ha de ser verdadera para que el conjunto (3.1) sea satisfacible. Por consiguiente, ninguna estructura puede hacer verdaderas simultáneamente a las fórmulas de (3.2).

El curso de este razonamiento puede organizarse en un diagrama como el siguiente:

10. Cf. Beth, 1955, y Hintikka, 1955. Para un estudio detallado de los cálculos analíticos y su relación con los cálculos de secuentes de Gentzen, cf. Smullyan, 1968. Una breve pero iluminadora explicación se da en el texto de Hodges al que se hace referencia en la nota 7.

$p \rightarrow (\neg q \wedge r),\ p \wedge \neg q,\ \neg r$

p

$\neg q$

$\neg p$ $\neg q \wedge r$

\# $\neg q$

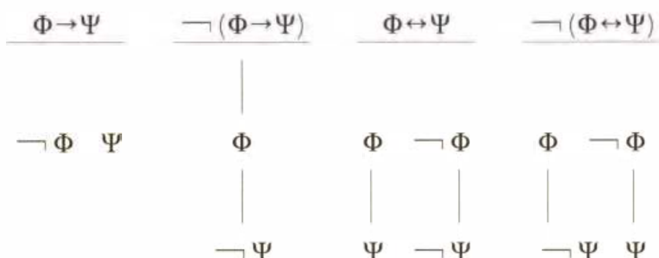
r

\#

En este tipo de diagrama, se parte de las fórmulas que se quiere hacer verdaderas. Cada estadio comporta una nueva exigencia para que ello sea posible. Cuando hay una ramificación, lo que tenemos es una alternativa: se requiere que las fórmulas de al menos una de las ramas sean verdaderas. Cuando, para alguna rama llegamos a la conclusión de que ello es imposible, puesto que en esa misma rama aparece —a lo largo de toda la ruta desde el origen— una fórmula simple y su negación, *cerramos* la rama (poniendo una marca, por ejemplo: #). Cuando todas las ramas están cerradas ello significa que es imposible hacer simultáneamente verdaderas a esas fórmulas. El conjunto de partida es insatisfacible, y la correspondiente pregunta sobre la implicación lógica se responde afirmativamente. Si, tras explorar todas las posibilidades, alguna rama quedara abierta, entonces ese conjunto es satisfacible, y la pregunta original sobre la implicación lógica habría que responderla negativamente.

Todas estas consideraciones pueden conducir a formular un cálculo puramente formal. En él los sucesivos estadios se alcanzan mediante la aplicación de reglas de derivación. Un conjunto de tales reglas puede ser el siguiente:

$\neg\neg\Phi$	$\Phi\wedge\Psi$	$\neg(\Phi\wedge\Psi)$	$\Phi\vee\Psi$	$\neg(\Phi\vee\Psi)$
Φ	Φ	$\neg\Phi$ $\neg\Psi$	Φ Ψ	$\neg\Phi$
	Ψ			$\neg\Psi$



Cada una de estas reglas expresa lo que se requiere para que la fórmula que encabeza la regla sea verdadera, como puede comprobarse por las tablas de verdad o por la definición de verdad (por ejemplo: para que $\neg(\Phi \leftrightarrow \Psi)$ sea verdadera se requiere o bien que Φ y $\neg \Psi$ sean verdaderas o bien que lo sean $\neg \Phi$ y Ψ). Pero estas consideraciones semánticas quedan fuera de la formulación estricta del cálculo.

Una de las cualidades de este cálculo es que suministra un *método de decisión* para la lógica de enunciados. Es decir, se puede utilizar el cálculo de una manera mecánica para determinar si una fórmula cualquiera del lenguaje es o no derivable, o derivable a partir de un conjunto de premisas dado (en la determinación de esto último son relevantes otras consideraciones que se harán más adelante, en la sección V).

Un cálculo formal sólo es un cálculo lógico si es *correcto,* es decir, si sólo pueden derivarse en él tautologías o implicaciones lógicas. Además, los cálculos lógicos son *completos* si todas las tautologías e implicaciones lógicas pueden derivarse en ellos. Los cálculos de la lógica de enunciados tienen estas dos propiedades. En cuanto a la corrección ello puede ser bastante obvio en algunos casos (como el de nuestro cálculo), pero la demostración de las afirmaciones anteriores cae por completo fuera del alcance de esta monografía. Más adelante, volveremos sobre estas y otras importantes propiedades (cf., de nuevo, la sección V).

IV. LOS LENGUAJES DE LA LÓGICA DE PRIMER ORDEN

Los lenguajes de primer orden disponen de lo esencial para hablar sobre los objetos de un *dominio* dado (también llamado *universo del discurso*), que puede cambiar contextualmente y sobre el que la única restricción que suele imponerse es que no sea vacío. «Hablar» quiere decir: predicar cosas sobre ellos, atribuirles propiedades, afirmar que están en determinadas relaciones o que no lo están, etc.; y ello bien sea en *aserciones particulares* —es decir, aserciones acerca de objetos determinados—, bien en *aserciones generales,* acerca de una multiplicidad de objetos —todos los del dominio, o alguno o algunos de ellos solamente.

Para la predicación de propiedades o relaciones se utilizan los *predicados* (incluyendo aquí expresiones de relación). Si las aserciones son par-

ticulares será necesario servirse además de las *constantes individuales,* que son el análogo de los nombres propios, y si son generales, se utilizan los *cuantificadores* y las *variables.*

Según esto, un lenguaje de primer orden tiene:

1) Variables: *x, y, z, u,* ...
2) Conectivas: $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$.
3) Cuantificadores: $\forall$, $\exists$.
4) Constantes individuales: *a, b, c,* ...
5) Predicados: *P, Q, R, S,* ...
6) Símbolo de identidad: $\approx$.
7) Paréntesis: (,).

Los símbolos de las tres primeras categorías son, con pequeñas salvedades, comunes a todos los lenguajes de primer orden. El número de variables puede variar, pudiendo ser finito o infinito numerable (tantas como números naturales: x_1, x_2, x_3, ...). Además, hay algunas variantes inesenciales respecto a las conectivas y los cuantificadores (conjuntamente conocidos como *símbolos lógicos*). Ya sabemos que no se limita el poder expresivo del lenguaje tomando algunos otros conjuntos alternativos. Tampoco se limitaría tomando un solo cuantificador como símbolo básico, pues cualquiera de ellos puede definirse en función del otro (con la ayuda de la negación).

El cuantificador $\forall$ es el *cuantificador universal.* Puede leerse «todos» o «todo». $\exists$ es el *cuantificador existencial;* pueden leerse «algún», «hay» o «al menos uno»[11].

Lo que varía de unos lenguajes a otros son las constantes individuales y los predicados. Es costumbre entonces utilizar las letras anteriores para hablar de un modo general sobre los lenguajes de primer orden y las propiedades lógicas de sus fórmulas, máxime cuando éstas no dependen del significado específico de los predicados.

Por último, el símbolo de identidad, un predicado especial que incluimos entre los símbolos lógicos, aumenta el poder expresivo del lenguaje[12].

Con el ejemplo siguiente se pretende ilustrar varios aspectos del funcionamiento de un lenguaje de primer orden. Supongamos que estamos hablando de ciertas propiedades de las personas de una determinada comunidad y de sus relaciones. Observamos, por ejemplo, que todos los miembros de la comunidad son varones. Esto lo expresaríamos así:

11. Existen variantes notacionales menos usadas, como lo son símbolos parecidos en la forma a las conectivas $\wedge$ y $\vee$, pero de mayor tamaño. Los cuantificadores siempre se escriben acompañados de variables y, aunque ahora ha caído en desuso, históricamente era frecuente escribir «*(x)*» para lo que hoy escribimos «$\forall x$».

12. La lógica de que hablaremos es, pues, la lógica de primer orden *con identidad,* extremo éste que a veces se especifica en la descripción. Las expresiones *lógica de predicados* y *lógica cuantificacional* se utilizan (con o sin la anterior especificación) esencialmente como sinónimas de *lógica de primer orden.*

(4.1) $\forall x \text{Varón}(x)$.

Algunos miembros son adultos, pero no todos:

(4.2) $\exists x \text{Adulto}(x) \land \neg \forall x \text{Adulto}(x)$.

Todos los adultos llevan barba.

(4.3) $\forall x(\text{Adulto}(x) \rightarrow \text{Barba}(x))$.

Algunos adultos son calvos, pero no todos los adultos lo son.

(4.4) $\exists x(\text{Adulto}(x) \land \text{Calvo}(x)) \land \neg \forall x(\text{Adulto}(x) \rightarrow \text{Calvo}(x))$.

Todos los adultos de la comunidad que llevan barba son tutor de alguien pero no todo tutor de alguien lleva barba.

(4.5) $\forall x((\text{Adulto}(x) \land \text{Barba}(x)) \rightarrow \exists y \text{Tutor}(x,y))$
$\land \neg \forall x(\exists y \text{Tutor}(x,y) \rightarrow \text{Barba}(x))$.

Todo tutor de un menor le enseña.

(4.6) $\forall x \forall y((\text{Tutor}(x,y) \land \text{Menor}(y)) \rightarrow \text{Enseña}(x,y))$.

Algún miembro de la comunidad tiene más de un tutor.

(4.7) $\exists x \exists y \exists z(\text{Tutor}(x,z) \land \text{Tutor}(y,z) \land \neg x \approx y)$.

Por último, hay una (única) persona que es el director de la comunidad.

(4.8) $\exists x(\text{Director}(x) \land \forall y(\text{Director}(y) \rightarrow x \approx y))$.

En las fórmulas anteriores, palabras como «Adulto», «Barba» y «Tutor» se han utilizado como predicados. Claro que no todos los predicados son sintácticamente del mismo tipo. Los dos primeros, por ejemplo, son *monádicos,* es decir, exigen solamente una variable o constante individual para formar una fórmula (algo análogo a una oración). El tercero es *diádico* (es decir, exige dos variables o constantes). Similarmente, en un lenguaje de primer orden puede haber predicados *triádicos, tetrádicos,* etc., es decir, predicados *n-ádicos,* para cualquier *n.*

Las fórmulas ilustran, de formas diversas, la interacción de los cuantificadores entre sí y de éstos con las conectivas, que como puede presumirse, puede llegar a ser muy compleja (no existe límite teórico para esta complejidad). En general, como vemos, «todos», «todo», se expresan mediante el cuantificador universal, mientras que «algunos», «alguien», «un» lo hacen mediante el existencial. Pero hay excepciones. Cuando hay una condición de unicidad y con «un» («uno», «una») se quiere decir «exactamente uno» (como en la última oración), la expresión en primer orden requiere además del cuantificador existencial, el universal y el signo de identidad. Y hay casos especiales que hacen traducir «un» mediante el cuantificador universal, como en (4.6)[13].

Todas las fórmulas ilustran el hecho ya señalado de que el dominio se determina contextualmente. En este caso el discurso es acerca de los miembros de una determinada comunidad, de manera que los valores que adoptan las variables son objetos o individuos de tal conjunto. Esta no ha sido una idea adoptada desde el origen de la lógica de primer orden.

13. El tipo de oración que traduce (4.6) se denomina a veces *oración «de burrito»,* debido a un ejemplo tradicional («Todo granjero que tiene un burro le pega»). Para los problemas semánticos que plantean, cf. Neale, 1990.

Frege, uno de los lógicos a quien más se debe el impulso de la lógica contemporánea, consideraba fijo el dominio de cuantificación, con lo cual se trivializaba éste, al tratarse del dominio de absolutamente todo. La diferencia puede explicarse en términos de nuestra ilustración tomando, por ejemplo, la fórmula (4.1). Mientras que para nosotros puede parafrasearse así: «Para todo individuo *de la comunidad,* ese individuo es varón», con la expresión «de la comunidad» dada por el contexto, para Frege habría que leerla de este otro modo: «Para todo individuo, ese individuo es varón». Si quisiéramos decir lo que decimos pero a la manera de Frege, con un solo universo de cuantificación, deberíamos utilizar una fórmula como:

(4.9) $\forall x(\text{Comunidad}(x) \rightarrow \text{Varón}(x))$,

donde la expresión «comunidad» expresaría la propiedad de ser un miembro de la comunidad en cuestión. Y algo análogo valdría para todos los demás casos.

Las dos últimas fórmulas permiten hacer ver el valor expresivo del símbolo de identidad, pues sin él en ninguno de los dos casos sería posible expresar la idea que se intenta expresar.

V. VERDAD E IMPLICACIÓN

La definición de las propiedades lógicas fundamentales para los lenguajes de primer orden sigue los mismos pasos que en la lógica de enunciados. Ello quiere decir que, ante todo, se debe caracterizar exactamente la sintaxis del lenguaje y el concepto de *verdad en una estructura,* para la noción de estructura adecuada al nuevo tipo de lenguajes.

La sintaxis delimita la noción de fórmula, mediante una definición recursiva como la de la sección I, sólo que ahora es algo más complicada. La nueva cláusula 1 dice que anteponiendo un predicado —del «número ádico» adecuado— al correspondiente número de variables o constantes (si el predicado es n-ádico, n constantes o variables) se obtiene una fórmula. Las cláusulas 2 y 3 quedarían igual en la nueva definición. Se añadiría ahora una cláusula al efecto de que anteponiendo a una fórmula un cuantificador acompañado de una variable, se obtiene también una fórmula. La nueva definición se cerraría con lo que en la anterior es la cláusula 4.

Es preciso hacer ahora una distinción que se revelará importante en lo que sigue: la que se da entre las fórmulas que contienen al menos una *variable libre* es decir, una variable que, al menos una de las veces que aparece no está afectada o *ligada* por ningún cuantificador (si x está en el alcance de $\forall x$ entonces está ligada por el cuantificador) y fórmulas que, o bien no contienen variables, o en las que todas las veces que éstas aparecen lo hacen ligadas por un cuantificador. A las primeras las llamaremos *fórmulas abiertas* y a las segundas *sentencias* (todos los ejemplos del apartado anterior lo eran de sentencias). Es posible formular esta distinción con mayor precisión, pero no es necesario hacerlo aquí.

Por lo dicho en la sección anterior, una estructura para un lenguaje de primer orden es un conjunto de objetos individuales (el dominio, que estipulamos no vacío), junto con algún «mecanismo» que asocie las constantes individuales del lenguaje con objetos del dominio y los predicados con propiedades adecuadas dado el «número ádico» de aquéllos (por ejemplo, si el predicado es diádico, con una relación binaria).

Ahora bien, lo que realmente se asocia a los predicados son *conjuntos adecuados, obtenidos a partir de los objetos del dominio.* A un predicado monádico, por ejemplo, se le asocia un subconjunto cualquiera (y uno sólo) de objetos del dominio; a uno diádico un conjunto de «díadas», es decir, de *pares ordenados* de objetos del dominio. Y así sucesivamente. El conjunto asociado a un predicado monádico contiene los objetos de los cuales es verdadero el predicado, a los que se aplica con verdad el predicado, como se reflejará más tarde en la definición de verdad para el lenguaje. El conjunto de pares asociado a un predicado diádico, contiene los pares de objetos del dominio a los que se aplica (en el orden que presenta el par) tal predicado. Etc.

Con ello, hablar de la propiedad o la relación asociada a un predicado queda sólo como una manera intuitiva de hablar que mantenemos cuando conviene.

Expresamos estas ideas de un modo más sucinto ayudándonos de la letra «*P*» que tomamos aquí para representar a cualquier predicado de *L,* el lenguaje de que se trate (ahora un lenguaje de primer orden)[14].

DEFINICIÓN 5.1. Una estructura $\mathbf{A}$ para *L* es un par $\langle$A, $\mathfrak{I}\rangle$ en el que A es un conjunto cualquiera no vacío e $\mathfrak{I}$ una función (llamada a veces *interpretación*) tal que:

1) $\mathfrak{I}$ asigna a cada constante individual de *L,* un elemento de A.

2) $\mathfrak{I}$ asigna a cada predicado *P* n-ádico de *L,* un conjunto de n-tuplos de elementos de A. ($n \geq 1$).

El próximo paso es definir la noción de *verdad en una estructura.* Se presenta aquí, sin embargo, una dificultad que no se daba en el caso de la lógica de enunciados. Allí se explicaba el modo en que el valor de verdad de un enunciado complejo dependía de los valores de verdad de sus constituyentes. El problema consiste en que ahora tenemos fórmulas complejas (las sentencias cuantificadas) cuyo valor veritativo no depende del de sus constituyentes, por la sencilla razón de que éstos, por el tipo de expresión que son, no tienen un valor veritativo. Por ejemplo, el valor veritativo de $\neg \Phi$ depende del de Φ, pero el de $\exists x P(x)$ no depende del supuesto valor veritativo de $P(x)$ pues esta fórmula, sencillamente, no tiene valor veritativo. Más intuitivamente: el valor veritativo de $\exists x$Varón(x) no depende del supuesto valor de Varón(x).

14. Utilizamos así esa letra de una manera ambigua, pues antes figuraba como un predicado de un lenguaje de primer orden, y ahora figura como una «letra esquemática» del metalenguaje. Pero el contexto resuelve bien estas ambigüedades y se evita así introducir aparato adicional.

Tarski, uno de los lógicos importantes del presente siglo, dio con la manera de salvar el escollo, introduciendo una nueva noción, la noción de *satisfacción,* y un recurso técnico[15]. Puede decirse que, esencialmente, todas los procedimientos que hoy se utilizan para hacerlo son variantes del suyo.

La noción de *satisfacción* es en principio muy sencilla. Consideremos una estructura $\mathbf{A} = \langle A, \mathfrak{I} \rangle$. Sea $\mathbf{a}_1$ un elemento de A (utilizamos las negritas como expresiones del *metalenguaje* para referirnos a elementos del dominio). Decimos que $\mathbf{a}_1$ satisface $P(x)$ si $\mathbf{a}_1$ es un elemento del conjunto $\mathfrak{I}(P)$ (el conjunto que $\mathfrak{I}$ asigna a P). Ahora bien, para el caso de un predicado diádico, digamos R, se necesitaría algo diferente: el *par* $\langle \mathbf{a}_1, \mathbf{a}_2 \rangle$ satisface $R(x,y)$ si ese *par* es un elemento del conjunto $\mathfrak{I}(R)$ (un conjunto de pares). Análogamente, habría que cambiar la definición para el caso de que el predicado fuera triádico, tetrádico, etc., por lo que, en rigor, se obtendría una familia de conceptos de *satisfacción* y no uno sólo. Además hay que considerar «casos mixtos», como la fórmula $R(x,b)$ con predicado diádico pero una sola variable libre.

Tarski solucionó el problema apelando a secuencias infinitas de objetos y a la noción de *satisfacción por una secuencia.* Pero su definición original no está hecha tomando en consideración la posibilidad de cambiar de dominio y de estructura. Por ello las variantes actuales son algo distintas. Aquí se utilizará la noción de *asignación* de objetos (de un dominio) a variables como noción auxiliar. Con esta noción resolvemos el problema original esencialmente con la observación de que, si bien una fórmula abierta como $P(x)$ no tiene en una estructura un valor veritativo «absoluto», sí lo tiene *relativamente* a la asignación de un objeto a la variable. Cuando se trate entonces de dar las condiciones de verdad de un enunciado como $\exists x P(x)$ simplemente se le declara verdadero si hay un objeto en el dominio que al ser asignado a la variable hace verdadera la fórmula. Esta es la idea básica que será desarrollada técnicamente a continuación.

Para ello necesitamos introducir nueva notación y algunas nociones auxiliares.

En la definición utilizamos las letras $t_1, \ldots, t_n$ para cualesquiera *términos individuales* (constantes individuales o variables). Nos servimos de la letra s para referirnos a una asignación de objetos del dominio a variables. Mediante $s(x/\mathbf{a})$ nos referimos a una asignación que es exactamente como s excepto posiblemente en que, sea lo que fuere lo que ésta asigna a la variable x, aquélla le asigna el objeto $\mathbf{a}$.

Es necesario un recurso para hacer referencia en general al objeto asignado a un término individual, sea éste una constante individual (caso en que la asignación la hace la función $\mathfrak{I}$ de la estructura) o a una variable (caso en que el objeto lo determina una asignación s). A tal fin introduci-

15. Cf. el famoso Tarski 1935, editado por vez primera en polaco en 1933.

mos una función s como extensión de s, de modo que $s(t) = s(t)$, si t es una variable, y $s(t) = \mathfrak{I}(t)$, si t es una constante individual. (El símbolo « = » es un símbolo del metalenguaje, a distinguir de « $\approx$ », el símbolo de identidad del lenguaje, aunque su función sea la misma).

La expresión «syss» abrevia a «si y sólo si». Utilizamos también la letra x para referirnos a cualquier variable (vale aquí también la observación de la nota 14). con Φ nos referimos ahora a una sentencia cualquiera del lenguaje de primer orden y con $\Phi(x)$ a cualquier fórmula que tenga una variable libre.

Finalmente, utilizamos $\| \Phi \|_{\mathbf{A},s} = \mathrm{V}$ para abreviar «Φ es verdadera en $\mathbf{A}$ con respecto a s», la noción que vamos a definir recursivamente.

Sea $\mathbf{A} = \langle \mathrm{A}, \mathfrak{I} \rangle$ una estructura para un lenguaje de primer orden L.

DEFINICIÓN 5.2.

1) Si P es un predicado n-ádico,
$\| P(t_1, \ldots, t_n)_{\mathbf{A},s} \| = \mathrm{V}$syss la n-tupla $\langle s(t_1), \ldots, s(t_n) \rangle$ es un elemento de $\mathfrak{I}(P)$.
2) $\| t_1 \approx t_2 \|_{\mathbf{A},s} = \mathrm{V}$ syss $s(t_1) = s(t_2)$.
3)-7) Igual que en la definición 2.1 pero con la relativización a s.
8) $\| \forall x \Phi(x) \|_{\mathbf{A},s} = \mathrm{V}$ syss, para todo elemento **a** de A,
$\| \Phi(x) \|_{\mathbf{A},s(x/\mathrm{a})} = \mathrm{V}$.
9) $\| \exists x \Phi(x) \|_{\mathbf{A},s} = \mathrm{V}$ syss, para al menos un elemento **a** de A,
$\| \Phi(x) \|_{\mathbf{A},s(x/\mathrm{a})} = \mathrm{V}$.

La noción de *modelo de* una sentencia o un conjunto de sentencias la podemos tomar ahora directamente de la sección II (teniendo en cuenta, claro está, la diferente noción de estructura subyacente) y llegamos así a la caracterización de las propiedades lógicas fundamentales para el caso de los lenguajes de primer orden. Sea L uno de estos lenguajes.

DEFINICIÓN 5.3.

a) $\Gamma \vDash \Phi$ syss todo modelo de Γ es también un modelo de Φ.
b) $\vDash \Phi$ syss toda estructura (para L) es un modelo de Φ.

Éstas son, para el caso de los lenguajes de primer orden, las nociones precisas de *implicación lógica* (o *consecuencia*) y, de *verdad lógica* o *validez,* como se dice en este caso (el término «tautología» no suele utilizarse fuera de la lógica de enunciados). Como puede comprobarse, son del todo análogas a las nociones correspondientes del caso más restringido de la lógica de enunciados. La diferencia radica en el tipo de estructura que es apropiada para uno y otro.

En el contexto más amplio de la lógica de primer orden también nos preguntamos por su capacidad como teoría lógica para dar cuenta de razonamientos intuitivamente válidos. En realidad es en el marco global de esta lógica, y no en el más limitado de la lógica de enunciados, cuando adquiere realmente sentido una respuesta. Pero en el planteamiento algo artificial «en dos tiempos» que aquí se hace, simplemente a las anterio-

res consideraciones añadimos lo que haya que decir acerca de los cuantificadores y oraciones cuantificadas.

También respecto a éstos es posible defender la lógica de primer orden en una línea similar a la que se esbozó en el caso de la lógica de enunciados. Puede decirse que las reglas semánticas para los cuantificadores (las correspondientes cláusulas de la definición de verdad) son una buena teoría de su contribución a las condiciones de verdad de las oraciones de las que forman parte, pues las condiciones de verdad de una oración cuantificada no dependen sino de los conjuntos de objetos a los que se aplican los predicados.

Una fuente de posibles objeciones concierne a la falta de presuposiciones existenciales del cuantificador universal en el análisis cuantificacional. $\forall x(P(x) \rightarrow Q(x))$ no implica $\exists x P(x)$; es más, la primera sentencia es verdadera si en el dominio no hay objetos a los que se aplique P. Ello origina en ciertos contextos algunos juicios aparentemente poco intuitivos sobre la verdad de las oraciones. Las consecuencias se ponen de manifiesto claramente en el rechazo por incorrectas de inferencias que en la lógica silogística tradicional se dan por correctas[16]. Por ejemplo, la inferencia (5.1) sería incorrecta.

(5.1) Todos los mamíferos son vivíparos
Todos los mamíferos tienen pulmones
―――――――――――――――――
Algún vivíparo tiene pulmones

Al juicio de incorrección se llega representando las dos premisas mediante sentencias cuantificadas universalmente como la del párrafo anterior y la conclusión mediante una sentencia existencial del tipo $\exists x(Q(x) \wedge R(x))$. Sea $\mathbf{A} = \langle A, \mathfrak{I} \rangle$ un modelo en el $\mathfrak{I}$ que asigna a P y a Q el conjunto vacío, siendo P y Q los representantes de «es un mamífero» y «es vivíparo», respectivamente. $\mathbf{A}$ es entonces un modelo de las premisas, pero no de la conclusión. Las premisas son verdaderas en esa estructura porque, como se desprende de la definición 5.2, una cuantificación universal de un condicional es verdadera *(vacuamente verdadera)* cuando el predicado del antecedente no se aplica a ningún objeto del dominio.

Las causas relevantes del juicio negativo sobre (5.1) son, por tanto, dos: nuestra decisión de representar las premisas como lo hemos hecho y el veredicto que emite la definición 5.2 sobre las cuantificaciones universales de condicionales con antecedente «vacuo». De tener serios motivos para defender la validez intuitiva de (5.1) habría que revisar una, al menos, de estas dos cosas. Revisar la segunda nos llevaría, desde luego, fuera del marco de la lógica cuantificacional clásica. Revisar la primera es mucho más fácil. Podríamos admitir que en el lenguaje cotidiano (al contrario de lo que sucede en el lenguaje matemático) afirmaciones uni-

16. Se rechazan los modos silogísticos *Darapti, Felapton, Bramantip* y *Fesapo*. Cf. la monografía sobre Historia de la lógica en el presente volumen.

versales como las de las premisas de (5.1) tienen también un cierto contenido existencial, de modo que habría que representarlas adecuadamente mediante una fórmula del tipo $\exists xP(x) \wedge \forall x(P(x) \rightarrow Q(x))$. Con ello se bloquearían contraejemplos como el anterior.

La cuestión es entonces si en el lenguaje cotidiano el significado de las oraciones en cuestión incluye el mencionado «contenido existencial». Para dilucidar esto sería aquí también relevante hacer la distinción entre las condiciones de verdad de una oración y los principios pragmáticos que rigen su uso (cf. nota 6).

VI. PROPIEDADES DE LA LÓGICA DE PRIMER ORDEN

Los cálculos de la lógica de primer orden no son sino ampliación de los de la lógica de enunciados, a los que basta con añadir unas pocas reglas, o axiomas (según de qué cálculos se trate) para las fórmulas con cuantificadores y el símbolo de identidad[17].

Sólo se dará aquí un ejemplo. Del cálculo analítico que se presentó en la sección 3 se obtiene un cálculo de la lógica de primer orden (dejando en este punto de lado la identidad) añadiendo las siguientes cuatro reglas[18]:

$$\frac{\forall x\Phi(x)}{\Phi(a)} \qquad \frac{\neg\, \exists x\Phi(x)}{\neg\, \Phi(a)} \qquad \frac{\exists x\Phi(x)}{\Phi(a)} \qquad \frac{\neg\, \forall x\Phi(x)}{\neg\, \Phi(a)}$$

Aquí, como anteriormente, la línea horizontal separa la fórmula a la que se aplica la regla de la fórmula resultado de su aplicación. Indica también que la aplicación de la regla no tiene por qué hacerse inmediatamente debajo de la fórmula a la que se aplica.

Las dos últimas reglas están sometidas a la restricción de que a sea una constante nueva, en el sentido de no utilizada aún en la derivación de que se trate.

Los cálculos ocuparon un papel central en la lógica clásica mientras las consideraciones semánticas se hacían sólo intuitivamente, al no haberse desarrollado aún las nociones semánticas de un modo sistemático, dentro de un marco matemático[19]. Hoy en día se ve su significación a través del *teorema de completud*. Por tanto, vamos a pasar ahora a exponer el contenido y el significado de este resultado y otros resultados importantes relacionados con él[20].

17. Todas las referencias de la nota 9 son pertinentes aquí.

18. $\Phi(a)$ es el resultado de sustituir en la fórmula $\Phi(x)$, con la x variable libre, esta variable por la constante individual a. En la formulación de la regla, la letra «a» figura como representante de cualquier constante.

19. Otro tanto puede decirse de las lógicas no clásicas. En las últimas décadas, la introducción de consideraciones semánticas sistemáticas ha revolucionado completamente el estudio de éstas. Cf. las monografías correspondientes de este mismo volumen.

20. La demostración de todos estos resultados exige el desarrollo del aparato de las pruebas por inducción y las nociones de teoría de conjuntos que las justifican. Estas pruebas pueden encontrarse en excelentes libros de texto como Boolos y Jeffrey, 1990; Enderton, 1972; Ebbinghaus, Flum y Thomas, 1984 y Smullyan, 1971.

Consideremos en adelante uno cualquiera de los cálculos lógicos a los que se ha aludido. Introducimos el signo «$\vdash$» para abreviar la afirmación de que algo es derivable en él. Así, $\Gamma \vdash \Phi$ significa que Φ es derivable a partir del conjunto de premisas Γ, y $\vdash \Phi$ significa que Φ es derivable sin premisas, que, como se dice, es un *teorema lógico.* (Este caso simplemente es el caso particular del anterior en el que Γ es el conjunto vacío, por lo que lo consideramos incluido en el caso general.)

La propiedad básica esencial para que un sistema de manipulación formal sea un cálculo lógico es la *corrección,* expresada en el siguiente teorema.

TEOREMA 6.1 *(Corrección):* Si $\Gamma \vdash \Phi$, entonces $\Gamma \vDash \Phi$.

Los cálculos de primer orden son *completos,* en el sentido del siguiente teorema[21], que expresa la primera propiedad interesante que esos cálculos poseen.

TEOREMA 6.2 *(Completud):* Si $\Gamma \vDash \Phi$, entonces $\Gamma \vdash \Phi$.

Una manera de explicar el significado profundo de este teorema es que muestra la medida en que es posible aproximarse al ideal leibniziano de sustituir la discusión racional por un procedimiento calculístico. En efecto, consideremos el conjunto de las verdades lógicas de un lenguaje de primer orden o de fragmentos del lenguaje natural representables en ese lenguaje. Es decir, el conjunto de las fórmulas u oraciones Φ de esos lenguajes tales que $\vDash \Phi$. Por el teorema de completud, existe una prueba formal de Φ en un cálculo de primer orden, es decir Φ es un teorema lógico, derivable sin premisas. Ahora bien, la cuestión de si algo constituye o no una prueba formal o deducción de una fórmula, es decidible mecánicamente. Pero esto implica que existe la posibilidad de un «listado mecánico» de tales teoremas: genérense mecánicamente todas las cadenas de símbolos y decídase mecánicamente cuáles son derivaciones de una fórmula; cada vez que se obtenga una que lo sea, añádase la fórmula a la lista.

La existencia de este procedimiento se expresa diciendo que el conjunto de los teoremas lógicos de primer orden es *recursivamente* (o *efectivamente*) *enumerable.* Pero, dada la corrección de los cálculos, esto significa que también el conjunto de verdades lógicas que considerábamos es recursivamente enumerable: es mecanizable la tarea de construir una lista tal que si una sentencia (de primer orden o representable en primer orden) es una verdad lógica, entonces aparecerá en la lista «en algún momento».

Esto no implica en absoluto que tengamos un procedimiento de *decisión,* en el sentido aludido arriba y brevemente descrito en la sección III. En este respecto la lógica de primer orden está en claro contraste con la de enunciados.

21. La primera prueba de este teorema se dio en Gödel, 1930. Las demostraciones que suelen darse en la actualidad proceden de Henkin, 1950.

El caso teóricamente más interesante, sin embargo, es el de las implicaciones lógicas. Ahora bien, cuando el conjunto de premisas es finito, este caso simplemente se reduce al anterior por (sucesivas aplicaciones de) el teorema 2.2 (que vale también en el contexto más amplio de la lógica de primer orden).

Piénsese que entre las inferencias que interesan se encuentran especialmente aquéllas en las que el conjunto de premisas es el conjunto de axiomas de una teoría matemática formulada en un lenguaje de primer orden. Los resultados anteriores revelan así la medida en que es posible mecanizar la prueba matemática[22], y son por ello de importancia para los fundamentos y la filosofía de la matemática. Y también son de significación para teorías de otras ramas de la ciencia, en tanto éstas puedan axiomatizarse en un lenguaje de primer orden o «contengan» teorías matemáticas.

Hemos explicado la relevancia del teorema de completud en el contexto de la *mecanización* de una prueba. La del teorema de *compacidad,* el siguiente de los importantes resultados lógicos de nuestra lista, ha de explicarse por su estrecha vinculación con la *existencia* de una prueba. Si $\Gamma \models \Phi$, el teorema de *completud* garantiza la existencia de una prueba formal (y, por lo tanto, de una prueba en definitiva), cualesquiera que sean Γ y Φ. Pero ¿cómo es esto posible en el caso de que Γ sea infinito si una prueba es necesariamente algo que ha de ser inspeccionable, constatable y, por tanto, finito? La respuesta se deriva del teorema de compacidad[23].

TEOREMA 6.3 *(Compacidad):* Si todo subconjunto finito de un conjunto de fórmulas es satisfacible, entonces este conjunto es también satisfacible.

La relevancia de este teorema se ve inmediatamente a través del siguiente teorema, llamado a veces teorema de *finitud,* que puede probarse con su ayuda[24].

TEOREMA 6.4 *(Finitud):* Si $\Gamma \models \Phi$, entonces hay un subconjunto finito Γ_0 de Γ tal que $\Gamma_0 \models \Phi$.

La propiedad de compacidad (o finitud) es así *necesaria* para la existencia de una prueba. Pero, juntamente con la enumerabilidad de los teoremas lógicos, es también *suficiente.* En efecto, la primera garantiza la

22. Para las teorías *finitamente axiomatizables* (axiomatizables con un conjunto finito de axiomas). La discusión se podría extender tomando en consideración las teorías cuyos axiomas, aunque infinitos, son enumerables recursivamente, pero no podemos aquí tomar este caso en consideración.
23. Probado por primera vez en Gödel, 1930.
24. De hecho, los dos teoremas son equivalentes, pues también es posible demostrar el de compacidad a partir del de finitud. La equivalencia de ambos teoremas hace que a veces se los formule como variantes el uno del otro.

reducción de las implicaciones lógicas al caso de un conjunto finito de premisas. Pero, como hemos visto, el teorema 2.2 reduce el caso al de averiguar la validez de una fórmula, y, si ésta es válida, el procedimiento de enumeración antes aludido proporcionaría su prueba.

Otro de los teoremas fundamentales, el llamado teorema de Löwenheim-Skolem[25], revela algo importante sobre las descripciones teóricas de un dominio.

TEOREMA 6.5 *(Löwenheim-Skolem):* Si Γ tiene un modelo, entonces tiene un modelo numerable.

El teorema de Löwenheim-Skolem fue en algún momento considerado paradójico. Así el propio Skolem se hizo la pregunta (llamada posteriormente *paradoja de Skolem)* de cómo era posible que un conjunto de fórmulas o proposiciones verdaderos sobre los números reales, que forman un conjunto no numerable, pudiera también ser verdadero de un ámbito numerable. Sin embargo, esta situación sólo puede verse como paradójica en el supuesto de que el objetivo de las teorías sea «definir» un ámbito determinado, es decir, caracterizar unívocamente un ámbito, de modo que no haya otro ámbito que responda a esa caracterización. El teorema nos señala lo injustificado de ese supuesto. Siempre que tengamos una teoría de una estructura infinita (en un lenguaje de primer orden), hay modelos de esa teoría «esencialmente distintos», es decir, no isomorfos (se dice entonces que la teoría en cuestión *no* es *categórica*). Lo mismo podemos decir, en particular, del método axiomático, que no es sino una forma distinguida de dar una teoría.

El teorema de compacidad y el de Löwenheim-Skolem describen propiedades «absolutas» de los lenguajes de primer orden, propiedades independientes de cualquier cálculo. Estas propiedades se utilizan para caracterizar el poder expresivo de los lenguajes de primer orden y de lenguajes que son *extensiones* de aquéllos[26]. Así, en la denominada *lógica* $\omega_1\omega$ se tienen fórmulas con un número infinito numerable de conjunciones o disyunciones, con lo que, por ejemplo, es posible expresar enunciados como «Todo objeto es un número natural» del siguiente modo:

(5.2) $\forall x(x \approx 0 \vee x \approx 1 \vee x \approx 2 \vee \ldots)$.

Surge así la cuestión de cuáles de esas extensiones «añaden realmente» algo a la capacidad expresiva de los lenguajes de primer orden. En este sentido no se añade nada si, por ejemplo, para cualquier sentencia de uno de esos lenguajes hay otra de primer orden que «dice lo mismo», es decir que es verdadera en las mismas estructruras. El teorema siguien-

25. Löwenheim (1915) ofreció una prueba del teorema para el caso de una sola fórmula y Skolem (1920) generalizó el resultado a conjuntos cualesquiera de fórmulas. Este teorema tiene variantes significativas; cf. Manzano, 1989.

26. Los lenguajes de segundo orden no tienen ninguna de las dos propiedades. Tampoco puede haber cálculos completos para ellos. Cf. la monografía sobre Lógica de orden superior en este volumen.

te[27], que presentamos en una formulación simplificada, caracteriza de una manera abstracta los lenguajes (y con ello la lógica) de primer orden.

TEOREMA 6.6 *(Teorema de Lindström).* Sea *L* una extensión del lenguaje de la lógica de primer orden para el que vale el teorema de Löwenheim-Skolem y el teorema de compacidad. Entonces toda sentencia de *L* tiene exactamente los mismos modelos que alguna sentencia de primer orden.

Las propiedades descritas en los teoremas anteriores son propiedades de gran importancia epistemológica (puesto que conciernen a lo que es posible demostrar y cómo), propiedades cuya presencia en la lógica de primer orden podemos probar precisamente por ser ésta una teoría matemática (en el sentido de «formulada matemáticamente», es decir, en el mismo sentido en que lo son las teorías de la física o la economía).

Hay otros lugares en que el hecho de que las teorías lógicas sean teorías matemáticas (en el sentido apuntado), tiene consecuencias dignas de ser notadas. Quizá la más prominente de todas sea que, al pasar de la presentación intuitiva de los conceptos a una teoría lógica, los dominios de que hablamos intuitivamente se precisan como *conjuntos,* y, como son precisamente las teorías matemáticas acerca de los conjuntos las que nos dicen lo que éstos son, las alternativas en esas teorías tienen repercusiones para la teoría lógica (cf. Jané, 1988-1989).

VII. VARIANTES DE LA LÓGICA DE PRIMER ORDEN

Filósofos contemporáneos tan importantes como Quine identifican la lógica de primer orden como *la* lógica sin más, proponiendo un lenguaje de primer orden —con la adecuada elección de predicados para satisfacer requisitos epistemológicos mínimos— como el *lenguaje canónico* de la ciencia[28]. Davidson hace que los lenguajes de primer orden jueguen un papel de condición cuasitranscendental de la inteligibilidad del discurso[29].

La defensa de tales afirmaciones depende de teorías filosóficas substanciales. Y la propia capacidad general de los lenguajes de primer orden para, con su ayuda, dar cuenta de diversos fenómenos semánticos y lógicas es controvertida (como se menciona en el apartado anterior). Lo que, sin embargo, parece menos dudoso es que la introducción de algunas variantes aumentaría la plausibilidad de ver en los lenguajes de primer orden buenos «modelos» de los lenguajes naturales, en el sentido que a esta palabra se le da frecuentemente en la ciencia empírica, es decir, ana-

27. Demostrado por vez primera en Lindström, 1969. Sobre la cuestión general de las propiedades y los límites de la lógica de primer orden, cf. Hodges, 1983.
28. Cf. Quine, 1960.
29. Cf. Davidson, 1967-1974.

logías simplificadas e idealizadas (no en el sentido técnico que tiene en secciones anteriores).

Versiones de la lógica de primer orden que no es infrecuente encontrar en los libros de texto incluyen los *símbolos funcionales* y, menos frecuentemente, un operador para expresar las *descripciones definidas.* Con ello se gana naturalidad en la representación del lenguaje natural y matemático en muchos contextos, sin obtenerse ninguna ampliación sustancial (al menos en el caso de los símbolos funcionales).

Las variantes de los lenguajes de primer orden que resultan de añadir símbolos funcionales y/o descripciones definidas no inciden, sin embargo, en el aspecto más central y seguramente más problemático, desde el punto de vista del presente tema, de los lenguajes de primer orden: la expresión de la cuantificación. Las dos variantes siguientes se dirigen precisamente a este aspecto central.

La primera obedece a una doble motivación, sintáctica y semántica. Desde el punto de vista sintáctico, tenemos la evidencia lingüística que apunta a la estructuración de la oración en sintagma nominal y sintagma verbal. Desde el semántico, tenemos la posibilidad de extender la capacidad expresiva de los lenguajes de primer orden introduciendo cuantificadores como «la mayoría», «más de la mitad», «ambos», «muchos», «pocos», «unos cuantos». Para ello habría que renunciar a la idea, debida a Frege, de que una afirmación cuantificada consiste en predicar algo, usualmente complejo, de todo o algún objeto del dominio de cuantificación. Así, «todo A es B» se analiza como una predicación compleja («si algo es A, entonces es B») de todo objeto, haciendo que el *término restrictivo,* A, pase a formar parte del predicado. Este rasgo es uno de los más característicos de los lenguajes de primer orden tal como se los formula en forma estándar y hace que, por ejemplo, «la mayoría de los A son B» no pueda expresarse en un lenguaje de primer orden, pues una afirmación de este tipo no puede reducirse a otra sobre la mayoría de los objetos de un dominio.

La nueva idea es construir un lenguaje de primer orden de manera que respete la unidad del determinante («todo», «algún», «el (la)», «muchos», etc. con el término restrictivo)[30]. Más precisamente, reemplazaríamos la cláusula para la introducción de los cuantificadores que se mencionaban al comienzo de la sección V por las dos cláusulas siguientes:

(7.1) Si $\Phi(x)$ es una fórmula donde x está libre, entonces $[\forall x{:}\Phi(x)]$ y $[\exists x{:}\Phi(x)]$ es un sintagma cuantificacional.

(7.2) Si $\Psi(x)$ es una fórmula donde x está libre, entonces $[\forall x{:}\Phi(x)](\Psi(x))$ y $[\exists x{:}\Phi(x)](\Psi(x))$ son fórmulas donde $[\forall x{:}\Phi(x)]$ y $[\exists x{:}\Phi(x)]$ son sintagmas cuantificacionales.

30. Cf. Barwise y Cooper, 1981 y Neale, 1990, para mayores detalles y para ver cómo se utiliza la nueva versión, respectivamente, en un análisis general de la cuantificación y en el de ciertos fenómenos anafóricos. El primero contiene además una prueba de la afirmación anterior acerca de «la mayoría».

Ejemplos similares a los que servían de ilustración en la sección IV serían los siguientes:

(7.3) Algunos adultos llevan barba.
$[\exists x\colon \text{Adulto}(x)]\ (\text{Barba}(x))$.

(7.4) Todos los menores tienen un tutor.
$[\forall x\colon \text{Menor}(x)]\ ([\exists y\colon \text{Cosa}(y)]\ (\text{Tutor}(y,x)))$;
$[\exists y\colon \text{Cosa}(y)]\ ([\forall x\colon \text{Menor}(x)]\ (\text{Tutor}(y,x)))$.

Hemos introducido el predicado monádico «Cosa» para ayudarnos en la representación de (7.4). Además, esta oración es ambigua, de modo que la primera fórmula vierte la interpretación en que los menores tienen tutores posiblemente distintos, mientras que la segunda corresponde a la interpretación en la cual el tutor es el mismo para todos ellos.

Si se quiere ahora extender el lenguaje de modo que incluya otros cuantificadores, ello se puede hacer reformulando las cláusulas anteriores de un modo muy sencillo.

También es sencilla la manera de modificar la definición 5.2 para obtener una definición de *verdad* para el nuevo lenguaje, restringiéndonos ahora de nuevo a los dos cuantificadores clásicos. Simplemente sustituiríamos las cláusulas 8 y 9 por las cláusulas siguientes:

(7.5) $\|\ [\forall x{:}\Phi(x)](\Psi(x))\ \|_{\mathbf{A},s} = V$ syss para todo elemento **a** de A tal que $\|\ [\Phi(x)\ \|_{\mathbf{A},s\,(x/a)} = V$, sucede $\|\ \Psi(x)\ \|_{\mathbf{A},s\,(x/a)} = V$.

(7.6) $\|\ [\exists x{:}\Phi(x)](\Psi(x))\ \|_{\mathbf{A},s} = V$ syss para al menos un elemento **a** de A tal que $\|\ \Phi(x)\ \|_{\mathbf{A},s\,(x/a)} = V$, sucede $\|\ \Psi(x)\ \|_{\mathbf{A},s\,(x/a)} = V$.

Veamos ahora para terminar otro tipo de modificación de la lógica de primer orden.

En muchos contextos, es natural suponer que los objetos de que se habla se agrupan en dominios distintos. Son de distinta «naturaleza», o como diremos, introduciendo el término técnico, son de *variedades* diferentes. Por ejemplo, supongamos que estamos hablando de las propiedades de ciertas familias y de sus miembros. Parece natural distinguir dos dominios o universos del discurso, uno constituido precisamente por familias y otro por personas o individuos.

Podemos «modelar» las afirmaciones de ese tipo mediante una extensión del lenguaje de primer orden. Distinguimos tipos distintos de variables, utilizando unas para hablar de objetos de una variedad y otras para objetos de otras variedades. Así, siguiendo con el ejemplo anterior, y utilizando las letras x, y,... para los individuos y α, β,... para las familias, podríamos hacer afirmaciones (posiblemente falsas) como las siguientes:

(7.1) Los hijos menores de una familia son inconformistas.
$\forall x(\exists \alpha\ \text{Menor}(x,\alpha) \rightarrow \neg\, \text{Conformista}(x))$.

(7.2) En toda familia de más de un hijo, de cada dos hijos siempre uno es «superior jerárquico» del otro.
$\forall \alpha \exists x \exists y((\neg\, x = y \wedge \text{Hijo}(x,\alpha) \wedge \text{Hijo}(y,\alpha)) \rightarrow (\text{Superior}(x,y) \vee \text{Superior}(y,x)))$.

En general, en la sintaxis de un *lenguaje multivariado* se distinguen distintas variedades de constantes individuales y de variables. Los predi-

cados del lenguaje sólo se combinan sintácticamente con n-adas de constantes o variables de las variedades apropiadas. Semánticamente, una estructura multivariada $\mathbf{A}$ consta de una serie de dominios $A_1, \ldots, A_m$, uno para cada variedad, sin elementos comunes, y una función $\mathfrak{I}$ que asigna a cada constante un objeto del dominio de la variedad correspondiente a esa constante y a cada predicado n-ádico un n-tuplo donde cada miembro es de la variedad apropiada.

Siguiendo las pautas descritas en secciones anteriores, se introduciría entonces el concepto de *verdad en una estructura multivariada, modelo de una sentencia* o *conjunto de sentencias de un lenguaje multivariado* y a continuación los de *sentencia válida* e *implicación lógica* para la lógica multivariada.

La cuestión ahora es si con ello tendríamos una teoría lógica diferente, quizás una ampliación de la lógica de primer orden en el mismo sentido que ésta amplía la de enunciados. La respuesta es que no, pues es posible reducir la lógica multivariada a lógica de primer orden, por el procedimiento que a continuación describimos en líneas generales.

En primer lugar, introducimos un lenguaje de primer orden L^* para el lenguaje multivariado L de que se trate, simplemente añadiendo a todos los símbolos de L una serie de predicados monádicos $P_1, \ldots, P_m$, uno para cada variedad. (En nuestro ejemplo, añadiríamos los predicados «Individuo» y «Familia».) A continuación establecemos una correspondencia entre las fórmulas de uno y otro de manera que a las fórmulas de las formas de (7.3) les hacemos corresponder, respectivamente, fórmulas de las formas de (7.4):

(7.3) $\forall w^i \Psi(w^i)$ y $\exists w^i \Psi(w^i)$

(7.4) $\forall x(P_i(x) \rightarrow \Psi(x))$ y $\exists x(P_i(x) \wedge \Psi(x))$,

donde w^i es una variable cualquiera de la variedad i, y x es una variable usual de primer orden. Llamemos Φ^* a la sentencia correspondiente a Φ. (La correspondiente a (7.1), por ejemplo, sería $\forall x$(Individuo$(x) \rightarrow$ ($\exists z$(Familia$(z) \wedge$ Menor$(x,z)) \rightarrow \neg$ Conformista$(x)))$.)

A continuación, convertimos cada estructura multivariada $\mathbf{A}$ para L en una estructura $\mathbf{A}^*$ para L^* tomando como nuevo dominio el conjunto de objetos que están en cualquiera de los dominios de $\mathbf{A}$ y con la misma asignación de objetos a constantes y predicados. Puede probarse entonces que una sentencia multivariada Φ es verdadera en $\mathbf{A}$ si y sólo si Φ^* es verdadera en $\mathbf{A}^*$. Consideremos ahora cualquier inferencia lógica $\Gamma \models \Phi$ de la lógica multivariada. Consideremos el conjunto Γ^* de sentencias correspondientes a las de Γ. Añadamos a éste un conjunto de fórmulas, Σ, en el que estén precisamente, para cada variedad i, las fórmulas $P_i(c)$ (para cada constante del lenguaje) y $\exists x P_i(x)$. El teorema siguiente nos da entonces la relación exacta que hay entre una inferencia en la lógica multivariada y una en la de primer orden.

TEOREMA 7.1. $\Gamma \models \Phi$ si y sólo si $\Gamma, \Sigma \models \Phi$.

En realidad basta con cumplir lo estipulado anteriormente para los predicados y constantes que aparezcan en la inferencia. Siguiendo con

nuestra ilustración, consideremos, por ejemplo, la siguiente afirmación de lógica multivariada:

$$\forall x(\exists\alpha \text{Menor}(x,\alpha) \rightarrow \neg \text{Conformista}(x)),\ \exists\alpha\text{Menor}(a,\alpha) \vDash \neg\text{Conformista}(a).$$

Por el teorema anterior, esta afirmación es verdadera si y sólo si la siguiente afirmación de la lógica de primer orden lo es:

$$\forall x(\text{Individuo}(x) \rightarrow (\exists y(\text{Familia}(y) \wedge \text{Menor}(x,y)) \rightarrow \neg \text{Conformista}(x))),$$
$$\exists y(\text{Familia}(y) \wedge \text{Menor}(a,y)) \vDash \neg \text{Conformista}(a).$$

Debido al teorema 7.1 podemos decir que la lógica multivariada no supone ninguna extensión esencial de la lógica de primer orden. No es necesario elaborar una teoría independiente de la lógica multivariada: las propiedades de la de primer orden (completud, compacidad, Löwenheim-Skolem) se transmiten a aquélla. Por otro lado, precisamente por esto mismo, podemos aprovechar cualesquiera ventajas en cuanto a intuitividad o naturalidad ofrezcan los lenguajes multivariados, sin perder ninguna de las que encontremos a la lógica de primer orden.

Los lenguajes multivariados tienen un interés teórico al permitir mostrar cómo el caso de la lógica de segundo orden con una semántica basada en las *estructuras generales* es reducible al de la lógica de primer orden. Simplemente, las variables de segundo orden toman valores en uno de los dominios de una estructura multivariada[31].

BIBLIOGRAFÍA

Anderson, J. M. y Johnstone (1962), *Natural Deduction: the Logical Basis of Axiom Systems,* Wadsworth, Belmont.

Barwise, J. y Cooper, R. (1981), «Generalized Quantifiers and Natural Languages»: *Linguistics and Philosophy, 4,* 159-219.

Beth, E. W. (1955), «Semantic Entailment and Formal Derivability», reimp. en Hintikka, 1969, 9-41.

Boole, G. (1847), *The Mathematical Analysis of Logic,* Macmillan, Barclay and Macmillan, Cambridge, recogido en Boole, G., *Studies in Logic and Probability,* Open Court, La Salle, 1952.

Boole, G. (1854), *An Investigation of the Laws of Thought* (publicado modernamente por Dover Publications, New York).

Boolos, G. S. y Jeffrey, R. C. ([3]1990), *Computability and Logic,* Cambridge University Press, Cambridge.

Davidson, D. (1967), «Truth and Meaning», en Davidson, 1984.

Davidson, D. (1974), «On the Very Idea of a Conceptual Scheme», en Davidson, 1984.

Davidson, D. (1984), *Enquiries into Truth and Interpretation,* OUP, Oxford, v.e. Gedisa, Barcelona, 1990.

Deaño, A. ([2]1978), *Introducción a la lógica formal,* Alianza, Madrid.

Ebbinghaus, H., Flum, J. y Thomas, W. (1984), *Mathematical Logic,* Springer Verlag, Berlin-Heidelberg-New York.

31. Cf. la monografía sobre Lógica de orden superior de este volumen, y también Enderton, 1972, 281-289.

Enderton, H. B. (1972), *A Mathematical Introduction to Logic,* Academic Press, New York.
Frege, G. (1879), *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens,* Halle, v.e. UNAM, México.
Frege, G. (1893), *Grundgesetze der Arithmetik, 1,* Jena.
Gabbay, D. y Guenthner, F. (comps.), *Handbook of Philosophical Logic,* 1, Reidel, Dordrecht.
Garrido, M. ([5]1981), *Logica simbólica,* Tecnos, Madrid.
Gentzen, G. (1934), «Untersuchungen über das logische Schliessen»: *Mathematische Zeitschrift, 39,* 176-210.
Gödel, K. (1930), «Die Vollständigkeit der Axiome des logischen Funktionenkalküls»: *Monatshefte für Mathematik und Physik,* 37, 349-360, v. ingl. Van Heijenoort, 1967; v.e. en Gödel, 1981.
Gödel, K. (1981), *Obras completas,* Alianza, Madrid. (Introducción y traducción de J. Mosterín.)
Grice, P. (1989), *Studies in the Way of Words,* Harvard University Press, Harvard.
Guttenplan, S. (1986), *The Languages of Logic,* Basil Blackwell, Oxford.
Henkin, L. (1950), «Completeness in the theory of types»: *Journal of Symbolic Logic, 15,* 81-91. (También en Hintikka, 1969.)
Hintikka, J. (1955), «Form and Content in Quantification Theory»: *Acta Philosophica Fennica,* 8, 7-55.
Hintikka, J. (1969) (comp.), *The Philosophy of Mathematics,* OUP, Oxford.
Hodges, W. (1983), «Elementary Predicate Logic», en Gabbay, D. y Guenthner, F. (comps.), 1983, 1-131.
Jané, I. (1988-89), «Lógica y ontología»: *Theoria, 10,* 81-106.
Kalish, D. y Montague, R. ([2]1980), *Logic: Techniques of Formal Reasoning,* Harcourt, Brace and World, New York.
Lindström, P. (1967), «On extensions of elementary logic»: *Theoria, 35,* 1-11.
Löwenheim, L. (1915), «Über Möglichkeiten im Relativkalkül»: *Mathematische Annalen, 76,* 447-470, v. ingl. en Van Heijenoort, 1967.
Manzano, M. (1989), *Teoría de modelos,* Alianza, Madrid.
Mates, B. ([4]1987), *Lógica matemática elemental,* Tecnos, Madrid.
Mosterín, J. ([2]1983), *Lógica de primer orden,* Ariel, Barcelona.
Neale, S. (1990), «Descriptive Pronouns and Donkey Anaphora»: *Journal of Philosophy, 87,* 113-150.
Peirce, C. S. (1902), «The Simplest Mathematics», en Peirce, 1933, vol. IV, 189-262.
Peirce, C. S. (1933), *Collected Papers of Charles Sanders Peirce,* C. Hartshorne y otros (comps.), Harvard University Press, Cambridge.
Quine, W. O. (1960), *Word and Object,* MIT Press, Cambridge, v.e. de M. Sacristán, Labor, Barcelona, 1968.
Quine, W. O. (1981), *Los métodos de la lógica,* Ariel, Barcelona.
Sacristán, M. ([2]1973), *Introducción a la lógica y al análisis formal,* Ariel, Barcelona.
Skolem, T. (1920), «Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen»: *Videnskapsselskapets skrifter, I. Matematisk-naturvidenskabelig klasse 3,* 1-36, v. ingl. en Van Heijenoort, 1967, 252-263.
Smullyan, R. (1968), *First Order Logic,* Springer-Verlag, Berlin-Heidelberg-New York.
Strawson, P. (1952), *Introduction to Logical Theory,* Methuen, London.
Sundholm, G. (1983), «Systems of Deduction», en Gabbay, D. y Guenthner, F. (comps.), 1983, 133-188.
Suppes, P. (1975), *Introducción a la lógica simbólica,* CECSA, México.
Tarski, A. (1935), «Der Wahrheitsbegriff in den formalisierten Sprachen»: *Studia Philosophica, 1,* 261-405, v. ingl. en Tarski, 1983.

Tarski, A. (1983), *Logic, Semantics, Metamathematics,* Hackett, Indianapolis.
Thomson, J. F. (1990), «In Defense of '⊃'»: *Journal of Philosophy, 87,* 57-70.
Van Heijenoort, J. (1967) (comp.), *From Frege to Gödel. A Source Book in Mathematical Logic, 1879-1931,* Harvard University Press, Harvard.
Whitehead, A. N. y Russell, B. (1910-1913), *Principia Mathematica,* 3 vols., Cambridge University Press, Cambridge.
Wittgenstein, L. (1921), *Tractatus Logico-Philosophicus,* en Annalen der Naturphilosophie, v. esp. J. Muñoz e I. Reguera, Alianza, Madrid, 1987.

LÓGICA DE ORDEN SUPERIOR

Ignacio Jané

I. INTRODUCCIÓN

Lo que distingue los lenguajes de primer orden de los segundo orden y, en general, de los de orden superior, es su restricción a cuantificar sólo sobre elementos del universo del discurso. Así, en el lenguaje de primer orden de la aritmética, es decir en el lenguaje de primer orden apropiado para hablar de números naturales con respecto al orden, la suma y el producto, podemos expresar propiedades de números (como la de ser par, primo, o perfecto), relaciones entre números (como la de divisibilidad o la de congruencia módulo 5) y podemos formular aseveraciones sobre números (como que hay un número infinito de primos, o que todo número natural es la suma de cuatro cuadrados) siempre que en las expresiones en cuestión nos limitemos a cuantificar sobre números naturales. Así, podemos expresar en primer orden que un número n es primo diciendo que *no hay ningún número* distinto de n y de 1 que divide a n; podemos expresar que un número n divide a m diciendo que *hay un número* cuyo producto con n es m; y podemos expresar que hay un número infinito de primos diciendo que *para todo número n hay un número m* que es primo y mayor que n. Pero en este lenguaje no podemos expresar hechos aparentemente tan simples como que *todo conjunto de números* naturales no vacío tiene un elemento mínimo; o que *toda propiedad* poseída por el cero y transmitida de un número a su sucesor (es decir: si un número la posee, su sucesor también) es poseída por todo número natural. En estas oraciones cuantificamos sobre conjuntos de números o sobre propiedades de números, no sobre números: cuantificamos no sobre elementos del universo del discurso, sino sobre subconjuntos de este universo o sobre propiedades de sus elementos. En lo que sigue, apenas hablaremos de propiedades, ya que todo cuanto podamos decir con su ayuda podremos reformularlo en términos de conjuntos. A

menudo (sobre todo en contextos matemáticos) lo único que importa de las propiedades es su extensión, y la extensión de una propiedad (razonable) es un conjunto.

Todos los lenguajes de que nos vamos a ocupar —ya sean de primer orden o de orden superior— serán lenguajes formales con símbolos de dos clases: lógicos y no lógicos (además de los símbolos impropios: los paréntesis). Los símbolos no lógicos de un lenguaje constituyen el *tipo de semejanza* del lenguaje en cuestión. Pueden ser de tres clases: constantes de predicado, constantes funcionales y constantes individuales. Las constantes funcionales y de predicado pueden serlo de distinto número de argumentos. Diremos que una constante funcional o de predicado de n argumentos ($n \geq 1$) es una constante n-aria. Así, el tipo de semejanza de un lenguaje apropiado para la aritmética constaría de:

— una constante de predicado binaria ($<$);
— dos constantes funcionales binarias ($+$ y $\times$);
— dos constantes individuales (0 y 1);

mientras que el tipo de semejanza del lenguaje usual de la teoría de conjuntos contiene un único símbolo no lógico: la constante de predicado binaria $\in$.

Los símbolos lógicos comunes a todos los lenguajes que aquí consideraremos (de todos los órdenes) son:

— los conectores ($\neg, \vee, \wedge, \rightarrow, \leftrightarrow$);
— el símbolo de igualdad ($\approx$);
— los cuantificadores ($\exists$, $\forall$).

Lo que, sintácticamente, distingue los lenguajes de órdenes distintos son las variables (que incluimos entre los símbolos lógicos). Los lenguajes de primer orden contienen una sola clase de variables: las *variables individuales:* $x_1, x_2, \ldots, x_k, \ldots$ Los lenguajes de segundo orden, además de las variables individuales contienen *variables de predicado:* para cada entero positivo n hay un número infinito de variables de predicado n-arias: $X^n_1, X^n_2, \ldots, X^n_k, \ldots$

Las variables toman valores. A cada clase de variable le corresponde un *dominio de variabilidad,* constituido por la totalidad de los valores que pueden tomar las variables de la clase. El dominio de variabilidad de las variables individuales es el universo del discurso. Los posibles valores de las variables de predicado unarias son subconjuntos de este universo, y los de las variables de predicado n-arias son relaciones n-arias en este universo. (Relaciones en extensión, se entiende, del mismo modo que los conjuntos son propiedades, o sea, relaciones unarias, en extensión. Una relación n-aria en un universo U es, pues, un conjunto de n-tuplas de elementos de U, es decir: un subconjunto de U^n).

Los lenguajes de orden superior a 2 (o los de orden infinito) contienen otras clases de variables. De ellos hablaremos más adelante. Ahora

es el momento de precisar los elementos sintácticos y semánticos de los lenguajes de segundo orden.

II. SINTAXIS DE LOS LENGUAJES DE SEGUNDO ORDEN

Supongamos que τ es un tipo de semejanza.

Los *términos* de cualquier *lenguaje sobre* τ son las *expresiones* (es decir, las sucesiones finitas de símbolos) caracterizados por las siguientes cláusulas recursivas:

1) Toda variable individual y toda constante individual es un término;
2) si f es una constante funcional n-aria y $t_1, \ldots, t_n$ son términos, entonces la expresión $ft_1 \ldots t_n$ es un término;
3) sólo son términos las expresiones obtenidas con ayuda de 1) y 2).

Una *fórmula atómica* del lenguaje de segundo orden sobre τ es una expresión de una de las tres formas siguientes:

$$t_1 \approx t_2, \qquad Rt_1 \ldots t_n, \qquad Xt_1 \ldots t_n,$$

donde $t_1, \ldots, t_n$ son términos, R es una constante de predicado n-aria de τ y X es una variable de predicado n-aria (así, $X = X^n_k$, para algún k).

Finalmente, las *fórmulas* del lenguaje de segundo orden sobre τ se caracterizan recursivamente así:

1) Toda fórmula atómica es una fórmula;
2) si α y β son fórmulas, también lo son $\neg\alpha$, $(\alpha \vee \beta)$, $(\alpha \wedge \beta)$, $(\alpha \rightarrow \beta)$ y $(\alpha \leftrightarrow \beta)$;
3) si α es una fórmula y x es una variable individual (así, $x = x_k$, para algún k), entonces $\exists x\alpha$ y $\forall x\alpha$ son fórmulas;
4) si α es una fórmula y X es una variable de predicado (así, $X = X^n_k$, para algún n y algún k), entonces $\exists X\alpha$ y $\forall X\alpha$ son fórmulas;
5) sólo son fórmulas las expresiones obtenidas con ayuda de 1) - 4).

Las variables, tanto individuales como de predicado, pueden aparecer libres o ligadas en una fórmula. No damos aquí la definición de estos términos, ya que es análoga a la correspondiente para lenguajes de primer orden. Baste recordar que una variable está ligada por un cuantificador en una fórmula si está en el alcance del cuantificador. Una fórmula con variables libres es una *fórmula abierta;* una fórmula sin variables libres es una *sentencia*.

III. SEMÁNTICA DE LOS LENGUAJES DE SEGUNDO ORDEN

Construimos lenguajes para hablar de objetos (los elementos del universo del discurso, muchas veces implícito) con respecto a ciertas relaciones

y operaciones entre ellos. Algunos de estos objetos, no necesariamente todos, tienen nombre en el lenguaje. Así, en aritmética hablamos de números con respecto a las operaciones de suma y producto y a la relación de orden; en la teoría de conjuntos hablamos de los conjuntos con respecto a la relación de pertenencia; en una posible teoría de las relaciones sociales de cierta sociedad hablaríamos de los individuos de esta sociedad con respecto a ciertas relaciones de parentesco; etc. El concepto general que engloba estos y otros muchos casos es el de estructura.

Si τ es un tipo de semejanza, una *estructura de tipo* τ es un par $\mathbf{A} = (A, \mathfrak{I})$, donde A es un conjunto no vacío —el *universo* de la estructura— e $\mathfrak{I}$ es una función (de *interpretación*) con dominio τ tal que:

1) Si R es una constante de predicado n-aria, $\mathfrak{I}(R)$ es una relación n-aria en A, es decir, un subconjunto de A^n;
2) si f es un símbolo funcional n-ario, $\mathfrak{I}(f)$ es una operación n-aria en A, es decir una función de A^n en A;
3) si c es una constante individual, $\mathfrak{I}(c)$ es un elemento de A.

Los términos sin variables denotan elementos del universo de la estructura y los términos con variables (necesariamente individuales) toman valores en la estructura dependiendo de qué elementos asignemos a estas variables. Así, si A es el conjunto de los números naturales, **A** es una estructura cuyo universo es A, $\mathfrak{I}(f)$ es la operación suma e $\mathfrak{I}(c)$ e $\mathfrak{I}(d)$ son los números 0 y 1, respectivamente, entonces el término *ffcdfdd* denota el número 3 $[=(0+1)+(1+1)]$; mientras que el término *ffxyfyz* toma el valor 12 $[=(2+3)+(3+4)]$, si asignamos los números 2, 3 y 4 a las variables x, y y z, respectivamente.

Las sentencias son verdaderas o falsas en una estructura, mientras que las fórmulas abiertas son satisfechas o no en la estructura por los objetos o relaciones que asignamos a sus variables libres. Un *modelo* de una sentencia es una estructura en la cual la sentencia es verdadera. Los conceptos de verdad y satisfacción para fórmulas de segundo orden son extensiones naturales de los correspondientes para fórmulas de primer orden, por lo que es innecesario definirlas con detalle (ver Enderton, 1972 o Boolos-Jeffrey, 1980). Basta tener presente que para determinar lo que una fórmula expresa en una estructura basta conocer la interpretación de los símbolos no lógicos, el significado de los conectores y cuantificadores, y recordar los dominios de variabilidad de las distintas clases de variables que en ella aparecen. Así, la fórmula $\forall x Xxx$, donde X es una variable binaria, es satisfecha en una estructura **A** si asignamos a X cualquier relación reflexiva en A, el universo de A, mientras que la fórmula $\forall x(Yx \rightarrow Zx)$, donde Y y Z son variables unarias, es satisfecha en **A** si asignamos a Y y a Z dos subconjuntos de A, el primero de los cuales está incluido en el segundo. Finalmente, la sentencia

INF: $\exists X(\forall x \neg Xxx \wedge \forall x \forall y \forall z((Xxy \wedge Xyz) \rightarrow Xxz) \wedge \forall x \exists y Xxy)$

es verdadera en una estructura **A** si y sólo si hay una relación binaria en A irreflexiva, transitiva y sin elementos maximales. Relaciones de este tipo las hay en conjuntos infinitos y sólo en ellos. Así: 1) es verdadera en una estructura si y sólo si su universo es infinito.

IV. ISOMORFISMO

Un *isomorfismo* entre dos estructuras $\mathbf{A} = (A, \mathfrak{I})$ y $\mathbf{A'} = (A', \mathfrak{I}')$ es una biyección h entre A y A' tal que:

1) Si R es una constante de predicado n-aria, $\mathfrak{I}(R) = \mathbf{R}$, $\mathfrak{I}(R') = \mathbf{R'}$; y $\mathbf{a}_1,\ldots, \mathbf{a}_n$ son elementos de A, entonces $R(\mathbf{a}_1,\ldots, \mathbf{a}_n)$ si y sólo si $R'(h(\mathbf{a}_1),\ldots, h(\mathbf{a}_n))$.
2) Si f es una constante funcional n-aria, $\mathfrak{I}(f) = \mathbf{f}$, $\mathfrak{I}(f') = \mathbf{f'}$ y $\mathbf{a}_1,\ldots, \mathbf{a}_n$ son elementos de A, entonces $h(\mathbf{f})$: $(\mathbf{a}_1,\ldots, \mathbf{a}_n)) = \mathbf{f'}(h(\mathbf{a}_1),\ldots, h(\mathbf{a}_n))$.
3) Si c es una constante individual e $\mathfrak{I}(c) = \mathbf{c}$, entonces $\mathfrak{I}'(c) = h(\mathbf{c})$.

Dos estructuras **A** y **A'** son isomorfas ($\mathbf{A} \equiv \mathbf{A'}$) si hay un isomorfismo entre ellas. Así, dos estructuras isomorfas sólo difieren, por así decir, en la naturaleza de los elementos de su universo. Por lo demás, toda descripción «formal» o «esquemática» o «estructural» de una de ellas lo es también de la otra. En particular, dos estructuras isomorfas son modelos de las mismas sentencias de primer y de segundo orden.

V. CAPACIDAD EXPRESIVA DE LOS LENGUAJES DE SEGUNDO ORDEN

La capacidad expresiva de los lenguajes de segundo orden es muy superior a los de primer orden. Hay muchos conceptos importantes que no son expresables mediante sentencias de primer orden pero sí de segundo orden. A continuación discutimos algunos de ellos. Unos nos permiten caracterizar clases de estructuras; otros definir ciertas relaciones u operaciones entre los elementos del universo de una estructura.

1. *Finitud e infinitud*

Consideremos el concepto de infinitud. Hemos visto que la sentencia INF de III es verdadera en las estructuras con universo infinito y sólo en ellas. Así, esta sentencia caracteriza la clase de las estructuras infinitas. El concepto de infinitud es, pues, expresable mediante una sentencia de segundo orden. Pero no lo es mediante una sentencia de primer orden.

En primer orden podemos expresar la infinitud del universo de la estructura de que hablamos mediante un conjunto infinito de sentencias:

$\Theta = \{\theta_2, \theta_3, \theta_4,\ldots \theta_n,\ldots\}$, donde
$\theta_2 = \exists x_1 \exists x_2 \neg x_1 \approx x_2$

$\theta_3 = \exists x_1 \exists x_2 \exists x_3(\neg x_1 \approx x_2 \wedge \neg x_1 \approx x_3 \wedge \neg x_2 \approx x_3)$

...

$\theta_n = \exists x_1 \exists x_2 \ldots \exists x_n(\neg x_1 \approx x_2 \wedge \neg x_1 \approx x_3 \wedge \ldots \wedge \neg x_{n-1} \approx x_n)$

...

pues, para cada n, la sentencia θ_n es verdadera en una estructura si y solo si su universo tiene por lo menos n elementos. Sin embargo, el concepto de finitud no puede expresarse en primer orden ni siquiera mediante un número infinito de sentencias. Un modo de verlo es el siguiente: Si suponemos que Φ es un conjunto de sentencias de primer orden verdaderas en todas las estructuras con universo finito y sólo en ellas, entonces todo subconjunto finito de $\Theta \cup \Phi$ tiene un modelo (pues todo subconjunto finito de Θ tiene un modelo finito). Así, por el teorema de compacidad de la lógica de primer orden, $\Theta \cup \Phi$ tiene un modelo: infinito, por serlo de Θ; finito, por serlo de Φ. Esto es absurdo. No existe, pues, un conjunto Φ tal.

Pero el concepto de finitud sí es expresable en segundo orden: mediante la negación de la sentencia INF.

FIN: $\neg$ INF

2. *Identidad*

La relación de identidad es definible mediante una fórmula de segundo orden, ya que el bicondicional

$$x \approx y \leftrightarrow \forall X(Xx \leftrightarrow Xy)$$

es satisfecho en una estructura por el par de objetos *a* y *b* si y sólo si *a* y *b* pertenecen a los mismos subconjuntos del universo de la estructura, lo cual ocurre si y sólo si *a* es igual a *b*. (Esta fórmula es una versión trivial del principio de la identidad de los indiscernibles: trivial porque debe su verdad al hecho de que para todo objeto, *a*, hay un conjunto cuyo único elemento es *a*.)

Pero la relación de identidad no es definible en primer orden. De ahí la necesidad de introducir en los lenguajes de primer orden el símbolo $\approx$ como símbolo lógico y, por tanto, no sujeto a reinterpretación; en segundo orden, como vemos, podemos prescindir de él.

3. *Transitivización de una relación*

La relación de antepasado a descendiente no es más que la transitivización de la relación de progenitor a hijo; es decir, es la menor relación transitiva que la extiende: *a* es un antepasado de *b* si y sólo si *a* es un progenitor de *b*, o un progenitor de un progenitor de *b*, o... Cuantificando sobre conjuntos de personas podemos definir la relación de antepasado a progenitor observando, con Frege, que ser un antepasado de *b* es ser miembro de todo conjunto de personas al que pertenecen los progenitores de *b* y los progenitores de todos sus miembros.

En general, si R es una relación, podemos definir en segundo orden la relación R^*, la transitivización de R, como sigue:

$$R^*xy \leftrightarrow \forall X((Rzy \rightarrow Xz) \wedge \forall u \forall w((Ruw \wedge Xw) \rightarrow Xu)) \rightarrow Xx).$$

4. *Operaciones aritméticas*

Con los medios de primer orden no podemos definir la relación que ordena el conjunto de los números naturales a partir de la relación R, que se da entre un número n y su sucesor inmediato: $R(n, m)$ si y sólo si $m = Sn$. Pero la relación que ordena los números naturales según su magnitud no es más que la transitivización de R: $n < m$ si y sólo si $m = Sn$, o hay k tal que $m = S(k)$ y $k = S(n)$, etc. Así: la relación de orden familiar entre los números naturales es definible en segundo orden a partir de la operación de sucesor:

$$x < y \leftrightarrow \forall X((XSx \wedge \forall u(Xu \rightarrow XSu)) \rightarrow Xy).$$

También, y en contraste con lo que ocurre en primer orden, son definibles en segundo orden las operaciones de suma y de producto a partir de la de sucesor. La relación ternaria R que subsiste entre los números n, m y k si y sólo si $n + m = k$ es la menor relación ternaria Z tal que (1) para todo número i: $Z(i,0,i)$, y (2) para cualesquiera números i, j, r: si $Z(i, j, r)$ entonces $Z(i, Sj, Sr)$. Así, si c es la constante cuya interpretación es el número cero:

$$x + y = z \leftrightarrow \forall Z((\forall u Zucu \wedge \forall u \forall v \forall w(Zuvw \rightarrow ZuSvSw)) \rightarrow Zxyz).$$

La definición del producto en términos de la suma (y, en definitiva, en términos de la operación de sucesor) es análoga. Basta observar que $n \times m = k$ si y sólo si n, m y k están relacionados por la menor relación ternaria Z tal que (1) para todo número i: $Z(i, 0, 0)$, y (2) para cualesquiera números i, j, r: si $Z(i, j, r)$ entonces $Z(i, Sj, r + i)$.

5. *Los números naturales*

Con sólo los medios disponibles en primer orden no podemos caracterizar el conjunto de los números naturales en términos de la operación de sucesor y del número cero. Ahora bien, los números naturales constituyen el menor conjunto que contiene el cero y también el sucesor de cada uno de sus miembros. Esto es expresable en segundo orden mediante el principio de inducción:

PI: $\forall X((Xc \wedge \forall x(Xx \rightarrow XSx)) \rightarrow \forall x Xx).$

Nos falta añadir las propiedades básicas del sucesor: (1) 0 no es sucesor de ningún número, y (2) la operación de sucesor es inyectiva. (1) y (2) son expresables en primer orden:

$$\forall x \neg Sx \approx c$$

$\forall x \forall y(Sx \approx Sy \rightarrow x \approx y)$.

Estas tres sentencias caracterizan la estructura de los números naturales con el cero y la operación de sucesor. Es decir, todo modelo (de tipo de semejanza $\tau = \{S, c\}$) de estas sentencias es isomorfo a ella.

6. *Numerabilidad*

Podemos caracterizar el orden de los números naturales directamente, sin reducirlo a la operación de sucesor: Es un orden lineal sin elemento máximo con respecto al cual todo conjunto acotado superiormente es finito. Dado que con los medios disponibles en segundo orden podemos expresar el concepto de finitud, podemos también caracterizar salvo isomorfismo la estructura de los números naturales con su orden.

Pero entonces también podemos expresar el concepto de numerabilidad. Un conjunto es numerable si es biyectable con el conjunto de los números naturales. Así, un conjunto es numerable si puede ordenarse con un orden isomorfo al de los números naturales, es decir, si hay un orden lineal en él con las propiedades recién enumeradas. Con algo más de detalle: Sea ORD(X) —donde X es una variable binaria— la fórmula que expresa que X es un orden lineal estricto (es decir, no reflexivo) del universo, y sea FIN(Y) —donde Y es unaria— la fórmula de segundo orden que expresa que toda relación irreflexiva y transitiva en Y posee un elemento maximal (de modo que Y es finito). Entonces la sentencia de segundo orden

NUM: $\exists X(\mathrm{ORD}(X) \wedge \forall x \exists y Xxy \wedge \forall Y(\exists z \forall y(Yy \rightarrow Xyz) \rightarrow \mathrm{FIN}(Y)))$

es verdadera en una estructura si y sólo si su universo es numerable.

7. *Continuidad*

El orden de los números reales, el continuo lineal, puede caracterizarse por las tres condiciones siguientes:

— Es un orden sin extremos;
— posee un subconjunto denso numerable;
— es condicionalmente completo.

Sabemos cómo expresar en primer orden que R es un orden sin extremos:

$\mathrm{ORD}(R) \wedge \forall x \exists y Rxy \wedge \forall x \exists y Ryx$.

Un subconjunto denso de (el universo de) un orden R es un conjunto X tal que entre dos elementos cualesquiera del orden hay por lo menos un elemento de X. En segundo orden sabemos expresar que un conjunto X es numerable mediante una fórmula, NUM(X), obtenida de modo análogo a la sentencia NUM. Así, también podemos expresar la segunda condición:

$\exists X(\mathrm{NUM}(X) \wedge \forall x \forall y(Rxy \rightarrow \exists z(Xz \wedge Rxz \wedge Rzy)))$.

Un orden es condicionalmente completo si todo subconjunto (de su universo) acotado superiormente posee una cota superior mínima. Que y es una cota superior mínima del conjunto X con respecto al orden R lo expresamos así:

$\mathrm{CSM}(y, X, R)$: $\forall x(Xx \rightarrow Rxy) \wedge \forall z(\forall x(Xx \rightarrow Rxz) \rightarrow (Ryz \vee y \approx z))$.

Así, la siguiente sentencia de segundo orden expresa que el orden R es condicionalmente completo:

$\mathrm{CC}(R)$: $\forall X(\exists y \forall x(Xx \rightarrow Rxy) \rightarrow \exists y(\mathrm{CSM}(y,X,R)))$

Por consiguiente, podemos expresar en segundo orden que R es un orden continuo, ya que los modelos de las tres sentencias que acabamos de escribir son precisamente los órdenes continuos lineales, es decir, los órdenes isomorfos al orden de los números reales.

Pero entonces también podemos expresar que un conjunto es biyectable con el conjunto de los números reales o, como suele decirse, tiene la cardinalidad del continuo. Sea $\alpha(Z)$ la fórmula obtenida formando la conjunción de las tres sentencias que expresan la continuidad del orden y reemplazando en ellas la constante R por la variable de predicado binaria Z. Así, $\alpha(Z)$ será satisfecha en una estructura **A** por una relación binaria en A si y sólo si ésta es un orden continuo de A. En consecuencia, la sentencia $\exists Z\alpha(Z)$ será verdadera en una estructura si y sólo si su universo tiene la cardinalidad del continuo.

8. *Biyectabilidad*

En segundo orden podemos también expresar que dos subconjuntos del universo de una estructura son biyectables. Una biyección entre X e Y es una función inyectiva de X sobre Y.
asociar a una función f una relación: la que subsiste entre dos objetos a y b si y sólo si $f(a) = b$. Así, dado que para expresar que una relación está asociada a una función y que esta función es una biyección entre X e Y es suficiente cuantificar sobre elementos del universo de la estructura, podemos expresar que X e Y son biyectables cuantificando sobre relaciones e individuos. Los recursos de segundo orden bastan para ello.

El truco aquí empleado de identificar funciones con ciertas relaciones es perfectamente general y nos permite concluir que todo cuanto podemos expresar cuantificando sobre funciones lo podemos expresar también cuantificando sobre relaciones y, por tanto, en segundo orden.

VI. LA RELACIÓN DE CONSECUENCIA

El concepto lógico fundamental es la relación de consecuencia: Una sentencia α es *consecuencia* de un conjunto de sentencias Σ si α es verdadera

en todo modelo de Σ, es decir, en todo modelo de todas las sentencias de Σ. Una propiedad fundamental de la relación de consecuencia en primer orden es su *carácter finito:* Si Σ es un conjunto de sentencias de primer orden y α es una sentencia de primer orden que es consecuencia de Σ, entonces α ya es consecuencia de algún subconjunto finito de Σ. Esta propiedad se pierde al considerar sentencias de segundo orden. Para verlo, basta recordar (ver V.1) que en segundo orden podemos expresar la infinitud del universo del discurso mediante una sola sentencia, INF, pero también (ver V.2) mediante un conjunto, necesariamente infinito, de sentencias de primer orden: $\Theta = \{\theta_2, \theta_3, \theta_4, \ldots, \theta_n, \ldots\}$. Así, INF es consecuencia de Θ. Pero no lo es de ningún subconjunto finito de Θ, ya que todo subconjunto finito de Θ tiene modelos finitos: si Γ es un subconjunto finito de Θ y n es el mayor subíndice de una sentencia de Γ, cualquier estructura cuyo universo tenga n elementos será un modelo de Γ.

Otro ejemplo de que la relación de consecuencia en segundo orden no es de carácter finito lo obtenemos considerando la relación de antepasado (ver V.3). Del conjunto infinito de sentencias que expresan que *a* no es un progenitor de *b,* ni un progenitor de un progenitor de *b,* ni un progenitor de un progenitor de un progenitor de *b,* ...se sigue que *a* no es un antepasado de *b*. Pero esto no se sigue, claramente, de ningún subconjunto finito suyo. En general, y con mayor precisión, si *R* es una constante de predicado binaria, *c* y *d* son constantes individuales, y Σ es el siguiente conjunto infinito de sentencias

$$\neg Rdc$$
$$\neg \exists x_1(Rdx_1 \wedge Rx_1c)$$
$$\neg \exists x_1 \exists x_2(Rdx_2 \wedge Rx_2x_1 \wedge Rx_1c)$$
...
$$\neg \exists x_1 \exists x_2 \ldots \exists x_{n-1} \exists x_n(Rdx_n \wedge Rx_nx_{n-1} \wedge \ldots \wedge Rx_2x_1 \wedge Rx_1c)$$
...

entonces $\neg R^*dc$, o, explícitamente, la sentencia

$$\neg \forall X((\forall z(Rzc \rightarrow Xz) \wedge \forall uw((Ruw \wedge Xw) \rightarrow Xu)) \rightarrow Xd)$$

es consecuencia de Σ, pero no de ningún subconjunto finito suyo.

La relación de consecuencia de un lenguaje es de carácter finito si y sólo si la lógica de este lenguaje es compacta. Que la lógica de un lenguaje sea compacta significa que siempre que todo subconjunto finito de un conjunto infinito de sentencias de este lenguaje tenga un modelo, el conjunto infinito también lo tendrá. La equivalencia entre la *compacidad* de una lógica y el carácter finito de su relación de consecuencia es fácilmente demostrable a partir de la observación según la cual una sentencia α es consecuencia de un conjunto de sentencias Σ si y sólo si el conjunto $\Sigma \cup \{\neg \alpha\}$, obtenido al añadir $\neg \alpha$ al conjunto Σ, no tiene modelo alguno.

Así, la lógica de primer orden es compacta, pero no la de segundo orden: todos los subconjuntos finitos de $\Theta \cup \{\text{FIN}\}$ poseen modelos, pero

$\Theta U\{FIN\}$ carece de ellos. Lo mismo ocurre con el conjunto $\Sigma U\{R^*$ dc$\}$.

Otra propiedad importante que distingue la lógica de primer orden de la de segundo orden es que en aquélla, pero no en ésta, valen los teoremas llamados «de Löwenheim-Skolem», que señalan limitaciones a la caracterización de estructuras con universo infinito. Así, según (una forma de) el *teorema descendente de Löwenheim-Skolem,* todo conjunto de sentencias de primer orden que posea modelos infinitos poseerá también un modelo numerable; mientras que (una versión débil de) el *teorema ascendente de Löwenheim-Skolem* dice que todo conjunto de sentencias de primer orden con modelos numerables tiene también modelos no numerables. La posibilidad de caracterizar la numerabilidad (V.6) y la continuidad (V.7) en segundo orden muestra que ninguno de estos dos teoremas es generalizable a la lógica de segundo orden.

VII. DEDUCIBILIDAD

La definición de la relación de consecuencia no es en absoluto constructiva. No sugiere ningún camino para determinar si una sentencia dada es consecuencia de un conjunto de sentencias, ni siquiera para obtener algunas de las consecuencias del conjunto. Si quisiéramos aplicar la definición con este fin, deberíamos tomar en consideración todas las estructuras y evaluar en ellas las sentencias en cuestión. Esto es, obviamente, impracticable.

Obtenemos consecuencias deduciendo, encadenando conclusiones obtenidas a partir de premisas mediante transformaciones formales, de índole sintáctica. Sistematizamos los métodos de deducción en cálculos deductivos. Los cálculos contienen reglas de inferencia, reglas que permiten obtener conclusiones de modo inmediato a partir de sus premisas. Tal vez contengan también axiomas, pero éstos son prescindibles, en cuanto un axioma puede considerarse como la conclusión de una regla sin premisas. Las reglas de inferencia del cálculo nos permiten construir deducciones. Dado un conjunto de sentencias Σ, una *deducción* a partir de Σ será típicamente una sucesión finita de fórmulas, cada una de las cuales o bien pertenece a Σ o bien se obtiene de las anteriores aplicando alguna regla de inferencia. Si α es la última fórmula de una deducción a partir de Σ decimos que α es *deducible* a partir de Σ (en el cálculo en cuestión).

No en todos los cálculos las deducciones son lineales o constan de fórmulas obtenidas exactamente del modo aquí descrito, pero en todos ellos las deducciones son finitas y, por tanto, sólo dependen de un conjunto finito de sentencias. La finitud de las deducciones es esencial si éstas deben servirnos para obtener fórmulas.

Un cálculo deductivo es *correcto con respecto a la consecuencia* si toda sentencia deducible a partir de un conjunto de sentencias es una conse-

cuencia del conjunto. Un cálculo es *completo con respecto a la consecuencia* si toda consecuencia de un conjunto de sentencias es deducible en el cálculo a partir del conjunto. Así, las sentencias deducibles a partir de un conjunto en un cálculo correcto y completo son exactamente las consecuencias del conjunto: en un cálculo tal, las relaciones de deducibilidad y de consecuencia coinciden extensionalmente.

No hay ningún cálculo deductivo correcto y completo con respecto a la consecuencia de los lenguajes de segundo orden. La razón es simple: Toda relación de deducibilidad es de carácter finito (pues si $\mathbf{d}$ es una deducción de α a partir de Σ, $\mathbf{d}$ es también una deducción de α a partir del conjunto Γ de sentencias de Σ que aparecen en $\mathbf{d}$ y, por la finitud de las deducciones, Γ es finito), pero, como sabemos, la relación de consecuencia no lo es. Así, deducibilidad y consecuencia son extensionalmente distintas.

Un caso límite de la relación de consecuencia lo constituye la verdad lógica, entendida como *validez universal.* Una sentencia α es universalmene válida si es válida en toda estructura. O, de modo equivalente, si es consecuencia de todo conjunto de sentencias. También creamos cálculos deductivos para obtener verdades lógicas. Un cálculo es *correcto con respecto a la validez* si toda sentencia deducible en él es universalmente válida, y es *completo con respecto a la validez* si toda sentencia universalmente válida es deducible en él. A partir de un cálculo correcto para la consecuencia obtenemos uno correcto para la validez teniendo en cuenta que una sentencia deducible sin premisas (es decir, deducible a partir del conjunto vacío de sentencias) es universalmente válida. Pero un cálculo puede ser correcto con respecto a la validez sin serlo con respecto a la consecuencia. Para verlo, basta observar que las reglas que permiten substituir en una sentencia una constante de predicado por otra del mismo número de argumentos son reglas que transforman verdades lógicas en verdades lógicas; pero su conclusión no es consecuencia de su única premisa, por lo que no pueden formar parte de ningún cálculo correcto con respecto a la consecuencia.

No hay ningún cálculo correcto y completo para la validez de las sentencias de segundo orden. La razón es más profunda que en el caso de la consecuencia. No depende únicamente de la finitud de las deducciones, sino de su efectividad. Para que un cálculo pueda ser usado para obtener fórmulas de cierto tipo (en este caso verdades lógicas) debe ser efectivamente decidible si una sentencia dada se obtiene como conclusión de una cierta regla de inferencia a partir de ciertas premisas. En tal caso, será posible decidir en un número finito de pasos si una sucesión finita de fórmulas es o no una deducción según las reglas del cálculo. Esto no significa que haya un método para decidir si una fórmula es o no deducible: sólo significa que habrá un método para determinar si una supuesta deducción de una fórmula lo es o no. Pero entonces habrá también un método efectivo para generar, una a una, posiblemente con repeticiones, todas las fórmulas deducibles en el cálculo. La idea es la siguiente:

puesto que sabemos generar efectivamente las sucesiones finitas de fórmulas y, de cada sucesión tal, sabemos decidir si es una deducción, podemos generar efectivamente las deducciones. Para generar las fórmulas deducibles, basta generar las deducciones y borrar de ellas todas las líneas menos la última.

En términos más precisos, el contenido del párrafo anterior es que el conjunto de las deducciones en un cálculo es recursivo, mientras que el de las fórmulas deducibles es recursivamente enumerable. Así, si hay un cálculo completo y correcto para la validez de segundo orden, el conjunto de las sentencias de segundo orden universalmente válidas será recursivamente enumerable. Pero, como veremos a continuación, no lo es.

En V.5 vimos que las tres sentencias

$$\forall x \neg Sx \approx c$$
$$\forall x \forall y (Sx \approx Sy \rightarrow x \approx y)$$
$$\forall X((Xc \wedge \forall x(Xx \rightarrow XSx)) \rightarrow \forall x Xx)$$

caracterizan la estructura $\mathbb{N}$ de los números naturales con el cero y el sucesor. Llamemos DED (por Dedekind) a su conjunción. Así, todo modelo de DED es isomorfo a $\mathbb{N}$, de modo que las consecuencias de DED son precisamente las sentencias de segundo orden (en el lenguaje cuyos símbolos son S y c) verdaderas en $\mathbb{N}$. Ahora bien, es claro que si α es una sentencia de este lenguaje, α es consecuencia de DED si y sólo si la sentencia DED$\rightarrow\alpha$ es universalmente válida. Por tanto: α es verdadera en $\mathbb{N}$ si y sólo si DED$\rightarrow\alpha$ es universalmente válida. De aquí se sigue que si hay un cálculo correcto y completo con respecto a la validez en segundo orden, entonces el conjunto de las sentencias de segundo orden verdaderas en $\mathbb{N}$ es decidible (recursivo). Un método para decidir si una sentencia α es verdadera en $\mathbb{N}$ es, a grandes rasgos, el siguiente: generamos una a una las sentencias universalmente válidas de este lenguaje hasta que damos con DED$\rightarrow\alpha$ o con DED$\rightarrow\neg\alpha$. Una de estas dos fórmulas debe aparecer, ya que o bien α o bien $\neg\alpha$ es verdadera en $\mathbb{N}$. Si aparece DED$\rightarrow\alpha$, α es verdadera en $\mathbb{N}$; si aparece DED$\rightarrow\neg\alpha$, $\neg\alpha$ lo es.

Sin embargo, el conjunto de las sentencias de segundo orden verdaderas en $\mathbb{N}$ no es decidible. Pues si lo fuera lo sería también el conjunto de las sentencias de primer orden verdaderas en la estructura $\mathbb{N}^*$ de los números naturales con la suma y el producto, que, por resultados de Church, Gödel y Tarski, no lo es (véase Tarski, Mostowski, Robinson, 1953, C.2). En efecto, en V.4 vimos cómo definir la suma y el producto en términos del cero y el sucesor mediante una fórmula de segundo orden. Podemos usar estas definiciones para eliminar los símbolos + y × de las fórmulas que las contengan, transformando de modo efectivo cada sentencia α del lenguaje de primer orden con los símbolos + y × en una sentencia α^* de segundo orden con sólo c y S, de modo que α sea verdadera en $\mathbb{N}^*$ si y sólo si α^* es verdadera en $\mathbb{N}$. Pero entonces, todo

método de decisión para el conjunto de las sentencias de segundo orden verdaderas en $\mathbb{N}$ da lugar a un método para decidir si una sentencia de primer (y también de segundo) orden es verdadera en $\mathbb{N}^*$. Dada α, constrúyase α^* y decídase si es verdadera en $\mathbb{N}$.

VIII. LENGUAJES DE ORDEN SUPERIOR FINITO

Supongamos que estamos interesados en la existencia de cierto tipo de colecciones de números naturales, por ejemplo, de cierto tipo de ideales. Un ideal sobre un conjunto A es un conjunto no vacío I de subconjuntos de A tal que, para cualesquiera subconjuntos X e Y de A, (1) si $X \in I$ e $Y \in I$, entonces $X \cup Y \in I$, y (2) si $X \in I$ y $Y \subseteq X$, entonces $Y \in I$. Para expresar, pues, que I es un ideal sobre A cuantificamos sobre subconjuntos de A; pero para decir que A admite cierta clase de ideales, es decir, que hay ideales sobre A con tales o cuales propiedades, debemos cuantificar sobre conjuntos de subconjuntos de A. Es un recurso que nos ofrecen los lenguajes de tercer orden.

En un lenguaje de tercer orden podemos cuantificar sobre individuos, es decir, sobre elementos del universo del discurso (como en primer orden); sobre conjuntos de individuos y relaciones entre ellos (como en segundo orden); pero además sobre conjuntos de conjuntos de individuos, sobre conjuntos de relaciones entre individuos, sobre relaciones entre relaciones entre individuos, etc. Sistematizando: Si A es el universo de la estructura sobre la que hablamos, sea $D_0 = A$; sea D_1 el conjunto de todos los subconjuntos de A; sea D_2 el conjunto de todas las relaciones binarias en A, y, en general, sea D_n el conjunto de las relaciones n-arias en A. Así, D_n es el dominio de variabilidad de las variables n-arias de segundo orden. Sea, finalmente, D la unión de los distintos D_n. Las variables de tercer orden toman como valores relaciones de cualquier número de argumentos entre elementos de D. Así, para cada sucesión finita de números naturales $n_1, \ldots, n_k$, un lenguaje de tercer orden contendrá, además de los recursos de segundo orden, variables k-arias de tipo $(n_1, \ldots, n_k)$, cuyos valores serán relaciones k-arias entre un elemento de D_{n1}, un elemento de $D_{n2}, \ldots$, y un elemento de D_{nk}.

De modo análogo, en los lenguajes de cuarto orden podemos cuantificar sobre conjuntos de conjuntos de conjuntos de elementos del universo del discurso A, y, en general, sobre relaciones de cualquier número de argumentos entre elementos de los distintos dominios de cuantificación de las variables de los lenguajes de tercer orden. Un lenguaje de orden $n+1$, en fin, se obtendrá de uno de orden n añadiéndole variables cuantificables cuyos valores serán relaciones entre elementos de los dominios de cuantificación de las variables del lenguaje de orden n.

IX. EL LENGUAJE DE LOS TIPOS

Todos los lenguajes hasta ahora considerados son fragmentos naturales del lenguaje de los tipos, un lenguaje con variables cuantificables de todos

los órdenes. Al igual que los anteriores, es un lenguaje apropiado para hablar sobre estructuras, de modo que, como ellos, más que un único lenguaje, lo que tenemos es una familia de lenguajes, uno para cada tipo de semejanza.

Empecemos por sistematizar el cúmulo de variables a nuestra disposición. Cada variable lo es de un cierto *tipo*, al igual que cada posible valor de una variable. El tipo básico es el de los individuos: 0; los demás tipos son sucesiones finitas de otros tipos. Los tipos se obtienen mediante las dos reglas siguientes:

0 es un tipo;

si $\eta_1,\ldots, \eta_n$ son tipos, la sucesión $(\eta_1,\ldots, \eta_n)$ es un tipo.

Cada tipo lo es de un cierto orden. El orden del tipo 0 es 1. El orden del tipo $(\eta_1,\ldots, \eta_n)$ es 1 + el máximo de los órdenes de los tipos $\eta_1,\ldots, \eta_n$.

Los objetos de tipo 0 son los individuos, es decir, los elementos del universo de la estructura en que interpretemos el lenguaje. 0 es el único tipo de primer orden (de orden 1). Los objetos de tipo $(\eta_1,\ldots, \eta_n)$ son las relaciones entre objetos de los tipos $\eta_1,\ldots, \eta_n$. Así (0) es el tipo de los conjuntos de individuos (0, 0) el de las relaciones binarias entre individuos, y, en general, (0, 0, ..., 0) (*n* ceros) es el tipo de las relaciones *n*-arias entre individuos. Éstos son los únicos tipos de segundo orden. ((0)), ((0), (0)), ((0), 0, 0) son tipos de tercer orden. El primero es el tipo de los conjuntos de individuos, el segundo el de las relaciones binarias entre conjuntos de individuos, y el tercero el de las relaciones ternarias entre un conjunto de individuos y dos individuos.

El lenguaje de los tipos contiene un número infinito de variables de cada tipo. El orden de una variable es el orden de su tipo. Las variables individuales son las variables de tipo 0. Los símbolos lógicos de este lenguaje son las variables, los conectores, los cuantificadores y el símbolo de igualdad. Los símbolos no lógicos son los del tipo de semejanza que consideremos.

Fijemos un tipo de semejanza τ. Los *términos de tipo 0* del lenguaje de los tipos sobre τ son los términos definidos en la sección 2. Si η es un tipo distinto de 0, los *términos de tipo η* no son más que las variables de tipo η. Las *fórmulas atómicas* del lenguaje de los tipos sobre τ son las expresiones de la forma

1) $t_1 \approx t_2$, donde t_1 y t_2 son términos de tipo 0.

2) $Rt_1\ldots t_n$, donde $t_1,\ldots, t_n$ son términos de tipo 0 y R es una constante de predicado *n*-aria de τ,

y las de la forma

3) $Xt_1\ldots t_n$, donde $t_1,\ldots, t_n$ son términos de los tipos $\eta_1,\ldots, \eta_n$, respectivamente, y X es una variable de tipo $(\eta_1,\ldots, \eta_n)$.

Las *fórmulas* se definen recursivamente como en los lenguajes de primer y de segundo orden: toda fórmula atómica es una fórmula; si α y β son fórmulas, también lo son $\neg\alpha$, $(\alpha\vee\beta)$, $(\alpha\wedge\beta)$, $(\alpha\to\beta)$ y $(\alpha\leftrightarrow\beta)$; si

α es una fórmula y X es una variable de cualquier tipo, $\exists X\alpha$ y $\forall X\alpha$ son fórmulas.

Para interpretar el lenguaje en una estructura **A**=(A, $\mathfrak{I}$), debemos especificar los dominios de variabilidad de los distintos tipos de variables. Los valores de las variables de tipo η serán objetos de tipo η sobre A. Así, para tipo η, definimos $D_\eta(A)$, el dominio de objetos de tipo η sobre A:

Si $\eta = 0$, $D_\eta(A) = A$;

Si $\eta = (\eta_1,\ldots, \eta_n)$, $D_\eta(A)$ es el conjunto de todas las relaciones *n*-arias entre elementos de $D_{\eta 1}(A),\ldots, D_{\eta n}(A)$, respectivamente.

Es ahora claro cómo especificar cuándo una fórmula es satisfecha en **A** por una asignación de objetos a sus variables libres (donde a una variable de tipo η se le asigna un elemento de $D_\eta(A)$). En el caso de las fórmulas atómicas de la forma (3), las únicas nuevas, la definición es la siguiente: La fórmula $Xt_1\ldots t_n$ es satisfecha en la estructura por una asignación de objetos a sus variables libres si los objetos asignados a $t_1,\ldots, t_n$ están en la relación asignada a la variable X. En cuanto a la cuantificación, basta recordar que si X es una variable de tipo η, $\exists X\ldots$ significa que hay un objeto de tipo η (un elemento de $D_\eta(A)$) tal que... Análogamente para $\forall X\ldots$ Los conceptos de consecuencia y validez universal pueden definirse ahora como antes.

La plétora de tipos puede reducirse grandemente gracias en parte a la posibilidad de identificar *n*-tuplas con conjuntos. Así, en el caso de los pares ordenados, podemos definir, como es usual en teoría de conjuntos,

$\langle x, y\rangle = \{\{x\},\{x, y\}\}$,

de modo que, si x e y son individuos, objetos de tipo 0, $\langle x, y\rangle$ es un conjunto de conjuntos de individuos: un objeto de tipo ((0)). En general, y con cierto cuidado, podemos reformular el lenguaje de los tipos usando sólo los tipos

0, (0), ((0)), (((0))), ...((...(0)...)), ...

Los dominios de variabilidad de las variables con respecto a una estructura cuyo universo es A son, entonces

$D_0(A) = A$
$D_{(0)}(A) = P(A)$
$D_{((0))}(A) = P(D_{(0)}(A)) = P^2(A)$
...
$D_{((\ldots(0)\ldots))}(A) = P(D_{(\ldots(0)\ldots)}(A)) = P^{n+1}(A)$
...,

donde $P(A) = \{X:X\subseteq A\}$, el conjunto potencia de A; y, para cada n, $P^{n+1}(A) = P(P^n(A))$, el conjunto potencia de $P^n(A)$.

X. LENGUAJE Y TEORÍA DE TIPOS

En el lenguaje de los tipos tratamos los conjuntos y las relaciones como objetos, de tal modo que ellos mismos pueden pertenecer a otros con-

juntos o estar relacionados mediante relaciones de orden superior. Pero usamos este lenguaje para hablar acerca de los elementos del universo de una estructura, no para estudiar los conjuntos y las relaciones en cuya base están estos elementos. Se trata de un lenguaje, no de una teoría de conjuntos o de relaciones. Qué conjuntos y qué relaciones hay en cada dominio se supone entendido al usar el lenguaje. Si queremos profundizar en el estudio del lenguaje, si queremos conocer mejor su semántica, debemos acudir a la teoría de conjuntos o, tal vez, a la *teoría* (en oposición al *lenguaje)* de los tipos. Una teoría tal nos iluminará acerca de lo que hay en cada dominio, nos dirá, por lo menos parcialmente, cómo obtener cada $D_\eta(A)$ a partir de A. Pero no lo olvidemos: el lenguaje que hemos introducido no es un lenguaje para hablar de conjuntos y relaciones sobre el universo de cierta estructura, sino para hablar, con ayuda de aquellos conjuntos y relaciones, sobre esta estructura. Si queremos estudiar conjuntos o relaciones acudiremos a una teoría apropiada que hablará de ellos. Ellos serán los individuos bajo observación, ellos constituirán el universo del discurso, el dominio de valores de las variables *individuales* del lenguaje que decidamos usar, el lenguaje en que formulemos esta teoría (que podría ser un lenguaje de tipos).

XI. CUANTIFICACIÓN PLURAL

En el lenguaje de los tipos y, en general, en los lenguajes con variables de orden tercero o superior, nos vemos obligados a tratar los conjuntos y las relaciones como objetos. No sólo los tratamos de hecho como objetos, sino que no podemos hacerlo de otro modo. Pues los valores de las variables de orden mayor que dos, digamos, para concretar, de las variables de tercer orden de tipo ((0)), son conjuntos cuyos elementos son a su vez conjuntos. Y los elementos de un conjunto son objetos. Así, el uso de un lenguaje con variables de tipo ((0)) nos obliga a dar cabida en nuestra ontología a conjuntos de individuos. En general, el uso de un lenguaje con variables de orden n + 2 comporta la reificación de conjuntos y relaciones de objetos de orden n, y el uso del lenguaje de los tipos presupone tratar como objetos a conjuntos de cualquier orden finito sobre el universo de la estructura que nos ocupe.

En segundo orden, la situación es distinta. Es cierto que, tal como hemos introducido la semántica, en los lenguajes de segundo orden cuantificamos sobre conjuntos de individuos y sobre relaciones entre ellos, de modo que, en tanto que elementos de un dominio de cuantificación, tratamos a unos y a otras como objetos. Pero, a diferencia de lo que ocurre en órdenes superiores, cabe la posibilidad de que la reificación de conjuntos y relaciones sea sólo un requisito de nuestro modo de presentar la semántica, ya que, en el lenguaje mismo, las variables de segundo orden sólo aparecen en posición de predicados y no en la de términos que requieren un predicado para constituir una fórmula; en otras palabras, no apa-

recen en posiciones reservadas a nombres o a pronombres: en posiciones propias de términos que denotan objetos. Los únicos términos de un lenguaje de segundo orden son las variables y las constantes individuales.

De hecho, es posible interpretar los lenguajes de segundo orden sin objetivizar conjuntos ni relaciones. El dominio de cuantificación es el mismo tanto si las variables cuantificadas son individuales como de predicado: el universo de la estructura considerada. La diferencia reside en el modo de cuantificar: cuantificación singular en el primer caso; cuantificación plural en el segundo (ver Boolos, 1984 y Lewis, 1991, 62-71). Es la diferencia que se ejemplifica en el par de oraciones: «Hay personas que se admiran sólo a sí mismas» y «Hay personas que se admiran sólo unas a otras». En el primer ejemplo, cada persona que sólo se admire a sí misma puede aducirse como prueba de la verdad de la oración; no así en el segundo, donde lo que se afirma es la existencia de varias personas con cierto comportamiento mutuo: cada una de estas personas, si admira a alguien, es a otra de ellas. La primera oración puede formalizarse en primer orden. Su forma es: $\exists x \forall y(Rxy \rightarrow x \approx y)$. Para la segunda, recurrimos a segundo orden: $\exists X(\exists x Xx \wedge \forall x \forall y((Xx \wedge Rxy) \rightarrow (\neg x \approx y \wedge Xy)))$. En la lectura usual, esta fórmula afirma la existencia de un conjunto no vacío de personas cada una de las cuales, si admira a alguien, es a otro miembro de este conjunto. Pero la lectura anterior, con cuantificación plural sobre personas, parece ser perfectamente apropiada. Cuando hablamos de conjuntos, o de grupos, de personas, no solemos pensar en el conjunto, en el grupo, como un objeto. Sólo se trata de un útil recurso de expresión.

Mediante la cuantificación plural, la definición de antepasado en términos de progenitor (ver V.3) sería: «*a* es un antepasado de *b* si y sólo si siempre que hay personas tales que 1.º) cada progenitor de *b* es una de ellas, y 2.º) si cada progenitor de cualquiera de ellas es una de ellas, *a* es también una de ellas». Tal vez esta formulación no sea más directamente inteligible que la de V.3, pero lo interesante es que es inteligible y que en ella se cuantifica sólo sobre personas, no sobre conjuntos de personas.

Hay dos puntos que es preciso mencionar acerca de la cuantificación plural como semántica de los lenguajes de segundo orden. El primero hace referencia al lugar del conjunto vacío. En la semántica usual, este conjunto es un valor de las variables de predicado, por lo que la cláusula «$\exists X$...» no debe, en rigor, interpretarse como «Hay objetos tales que...» sino más bien como «Hay cero o más objetos tales que...». Si no lo hacemos, ciertas fórmulas, como $\exists X \forall y(Xy \leftrightarrow (Py \wedge \neg Py))$, toman distinto valor de verdad en las dos semánticas. El segundo punto tiene que ver con la interpretación de las variables de predicado de más de un argumento. Si la cuantificación de variables unarias se interpreta como cuantificación plural sobre los elementos del dominio, la cuantificación de variables binarias debería interpretarse como cuantificación plural sobre pares de elementos del dominio, la de variables ternarias sobre triples de elementos del dominio, etc. Así, si X es binaria, las sentencias

$\exists X \forall x Xxx$ $\quad\quad\quad$ $\exists X \forall x \forall y (Xxy \rightarrow Xyx)$

podrían leerse, respectivamente, como: «Hay pares de objetos tales que para todo objeto *x* el par *xx* es uno de ellos» y «Hay pares de objetos tales que para cualesquiera objetos *x, y:* si el par *xy* es uno de ellos, también lo es el par *yx*». Admitir pares, triples, y, en general, *n*-tuplas de elementos de un dominio no es ontológicamente mucho más comprometido que admitir los elementos mismos, ya que siempre podemos precisar cuáles son estos pares, estos triples, etc., en términos de los elementos del dominio. Este no es el caso con los conjuntos de elementos de un dominio infinito. De ahí que la cuantificación plural, incluso sobre pares, triples y *n*-tuplas, pueda verse como un intento de descargar ontológicamente la lógica de segundo orden que, según el conocido aforismo de Quine (1970, 66), es teoría de conjuntos con piel de cordero.

XII. ESTRUCTURAS GENERALES

La gran capacidad expresiva de la lógica de segundo orden y superior se debe a la riqueza de sus dominios de cuantificación. Así, si en segundo orden podemos caracterizar el orden de los números reales es porque nuestras variables toman como valores todos los subconjuntos del universo de la estructura que consideremos. Si por ejemplo la relación de orden de la estructura considerada, digamos (A, $\mathfrak{I}$), no es condicionalmente completa, habrá un subconjunto acotado de A sin cota superior mínima. Pero este subconjunto, llamémosle S, será un valor de las variables unarias. S satisfará la fórmula (ver V.7)

1) $\exists y \forall x (Xx \rightarrow Rxy)$,

pero no satisfará la fórmula

2) $\exists y (\mathrm{CSM}(y, X, R))$

y, por tanto, en nuestra estructura será falsa la sentencia CC(R), que expresa que $\mathfrak{I}(R)$ es un orden condicionalmente completo.

(Lo mismo ocurriría si interpretáramos el lenguaje de segundo orden según la cuantificación plural, si bien la fraseología sería algo menos perspicua: Si S es un conjunto acotado sin cota superior mínima, sus elementos son testigos de la falsedad de CC(R), ya que 1) son números tales que hay un número mayor o igual que todos ellos, pero 2) no hay ningún número mayor o igual que todos ellos que sea menor o igual que todos los números mayores o iguales que todos ellos).

Como vimos en X, la riqueza de los distintos dominios de cuantificación se presupone en el uso de los lenguajes de segundo orden y de orden superior. Esto significa que si usamos un lenguaje de segundo orden para comunicarnos, los interlocutores debemos entender del mismo modo la referencia a «todo subconjunto» y a «toda relación» (ver Shapiro,

1985). Si no fuera así, ¿cómo podríamos concluir que, continuando con el ejemplo anterior, el orden que nuestro interlocutor dice satisfacer la sentencia CC(*R*) es condicionalmente completo? No podríamos. Si el conjunto S antes citado no es uno de los valores posibles de sus variables de predicado unarias (porque no está en *su* dominio de cuantificación pertinente), nuestro interlocutor podrá tomar por condicionalmente completo un orden que no lo es.

Esta presuposición puede resultar irrazonable. ¿Cómo podemos garantizar que, por ejemplo, nuestro dominio de cuantificación $D_{(0)}(A)$ contiene *todos* los subconjuntos de A si, en el caso en que A sea infinito, la mayor parte de ellos serán totalmente indescriptibles? Ciertamente sabemos reconocer un subconjunto de A si nos es dado (descrito), también sabemos qué significa ser un subconjunto de A: ser un conjunto cuyos elementos pertenecen todos a A; pero de que sepamos estas y otras cosas no se sigue que tengamos una concepción precisa de qué conjuntos hay en P(A), es decir, de la totalidad de los subconjuntos de A. De ahí que se haya considerado una semántica alternativa para los lenguajes de orden superior, una semántica que no presuponga la comprensión unívoca de «todo predicado», «toda relación», etc.

Presentemos esta semántica, debida a Leon Henkin (1950), para el lenguaje de los tipos. Si **A** = (A, $\mathfrak{I}$) es una estructura, una *estructura general* sobre **A** consta de **A** y de una función, C, cuyo dominio es el conjunto de los tipos, tal que (poniendo 'C_η' en vez de '$C_{(\eta)}$'):

Si $\eta = 0$, $C_\eta(A) = A$;
Si $\eta = (\eta_1, \ldots, \eta_n)$, $C_\eta(A)$ es *un* conjunto de relaciones *n*-arias entre elementos de $C_{\eta 1}(A), \ldots, C_{\eta n}(A)$, respectivamente

En una estructura general sobre **A** podemos interpretar el lenguaje de los tipos como hicimos en IX, con la única diferencia que ahora las variables de tipo η toman como valores elementos de $C_\eta(A)$. $C_\eta(A)$ es, pues, el dominio de cuantificación de tipo η de la estructura general. Es claro que cada $C_\eta(A)$ es un subconjunto de $D_\eta(A)$. La estructura general máxima, aquella en la cual para cada η: $C_\eta(A) = D_\eta(A)$, es la *estructura principal* sobre **A**.

Es conveniente limitarse a considerar estructuras generales en las que sean verdaderos todos los *axiomas de comprehensión:*

$$\forall X_1 \ldots \forall X_n \exists Y \forall Z_1 \ldots \forall Z_n (Y Z_1 \ldots Z_n \leftrightarrow \alpha),$$

donde $X_1, \ldots, X_n$ son variables de cualquier tipo, $Z_1, \ldots, Z_n$ son variables de tipos $\eta_1, \ldots, \eta_n$, respectivamente, Y es una variable de tipo $(\eta_1, \ldots, \eta_n)$, y α es una fórmula todas cuyas variables libres se encuentran entre $X_1, \ldots, X_n, Z_1, \ldots, Z_n$.

Estos axiomas garantizan que, para cada tipo η, todas las relaciones definibles de tipo η pertenecerán al dominio de cuantificación $C_\eta(A)$. En particular, el complemento, la unión y la intersección de elementos de

$C_\eta(A)$ serán a su vez elementos de $C_\eta(A)$. Esto es algo que esperamos de un dominio de cuantificación razonable: Si $S_1,\ldots, S_n$ son posibles valores de las variables, también debe serlo toda relación definible en el lenguaje a partir de $S_1,\ldots, S_n$.

Una estructura general sobre **A** puede ser muy distinta de la correspondiente estructura principal. Así, si A es numerable, hay estructuras generales sobre **A** tales que cada dominio de cuantificación $C_\eta(A)$ es también numerable. Pero la cardinalidad de los dominios de cuantificación de la estructura principal sobre **A** crece exponencialmente con el orden de los tipos. $D_0(A)$ tiene cardinalidad $\aleph_0$; $D_{(0)}(A)$ tiene la cardinalidad del continuo: $\mathfrak{c} = 2^{\aleph_0}$; $D_{((0))}(A)$ tiene cardinalidad $2^{\mathfrak{c}}$; etc. (La enorme cardinalidad de los dominios de cuantificación en la interpretación principal del lenguaje de los tipos puede aducirse en contra de la razonabilidad de la semántica usual.)

Con la semántica de las estructuras generales, el lenguaje de los tipos se convierte esencialmente en un lenguaje de primer orden. Definamos los conceptos lógicos de esta semántica: una sentencia es consecuencia de un conjunto de sentencias Σ si es verdadera en todos los modelos generales de Σ; y una sentencia es universalmente válida si es verdadera en todas las estructuras generales. Para esta relación de consecuencia en sentido general hay un cálculo correcto y completo, y, por tanto, hay también un cálculo correcto y completo para la validez lógica (Henkin, 1950). Por la existencia de un cálculo completo y correcto para la consecuencia, esta relación es de carácter finito y, por consiguiente, vale el teorema de compacidad. También los teoremas de Löwenheim-Skolem se cumplen: Si un conjunto finito o numerable de sentencias tiene un modelo general con universo infinito, tiene también un modelo con universo infinito de cualquier cardinalidad.

Pero el gran poder expresivo de los lenguajes de orden superior se desvanece al darles la semántica de las estructuras generales. No podemos caracterizar la finitud: la sentencia $\neg$ INF de V.1 tiene modelos generales infinitos; no podemos definir la transitivización de una relación (ver V.3): si el conjunto de los antepasados de *a* no está en $C_{(0)}(A)$, *b* puede pertenecer a todos los conjuntos de $C_{(0)}(A)$ que contengan a los progenitores de *a* y a los progenitores de todos sus miembros sin ser por ello un antepasado de *a;* no podemos tampoco caracterizar los números naturales (V.5) ni el orden de los números reales (V.7), etc. Ni siquiera podemos definir la relación de identidad (V.2): si no introducimos el símbolo de igualdad como un símbolo lógico del lenguaje (interpretándolo como identidad) sino que lo introducimos mediante la definición

$$x \approx y \leftrightarrow \forall X(Xx \leftrightarrow Xy),$$

podemos mostrar sin gran dificultad que hay estructuras generales que contienen elementos distintos que, sin embargo, satisfacen la fórmula $x \approx y$ porque no hay suficientes conjuntos en $C_{(0)}(A)$ para distinguirlos.

Así, la semántica de las estructuras generales desvirtúa la lógica de orden superior, rebajándola en cierto sentido a primer orden. La razón de que esto ocurra es, a grandes rasgos, la siguiente: Como hemos visto, en esta semántica los dominios de cuantificación superiores están sujetos a interpretación; no están, como en la semántica usual, determinados por el universo individual de la estructura. La libertad en la elección de estos dominios de cuantificación es tanta que la totalidad de las estructuras generales sobre estructuras de un tipo de semejanza τ puede describirse, salvo isomorfismo, mediante un conjunto de sentencias de primer orden en un tipo de semejanza ampliado τ^*, obtenido al añadir a τ constantes de predicado que se interpretarán como los distintos universos de cuantificación. De este modo, hacer lógica de orden superior sobre estructuras de tipo de semejanza τ es equivalente a hacer lógica de primer orden sobre estructuras de tipo de semejanza τ^* (ver Enderton, 1972, C. 4).

XIII. CONSIDERACIONES FINALES

Dejemos atrás la semántica de las estructuras generales y volvamos a la semántica usual, la que da a los lenguajes de orden superior su gran potencia expresiva. Permanezcamos, además, en segundo orden, el lenguaje que, en parte por razones de simplicidad, hemos descrito con mayor detalle. Restringirnos a segundo orden no es una decisión arbitraria, ya que el gran salto en la capacidad de caracterización de estructuras y de clases de estructuras se da entre primer y segundo orden. En cierto modo, los lenguajes de orden mayor que dos pueden reducirse a los de segundo orden (ver Van Benthem, Doets, 1983, 323-324).

Una diferencia fundamental entre los lenguajes de primero y los de segundo orden tiene que ver con la evaluación de una sentencia en una estructura. Si α es una sentencia de primer orden, la verdad o falsedad de α en una estructura **A** depende únicamente de α y de lo que está explícitamente dado al dar **A**, a saber: su universo, A, y la función de interpretación, $\mathfrak{I}$, de los símbolos del lenguaje. Nada más es necesario. La evaluación de α en **A** puede llevarse a cabo con sólo estos ingredientes, pues 1.º el valor de una fórmula atómica depende únicamente de la interpretación de las constantes que en ella aparecen y de los elementos de A asignados a sus variables; 2.º el valor de una fórmula obtenida a partir de otras con ayuda de los conectores ($\neg$, $\vee$, $\wedge$. $\rightarrow$, $\leftrightarrow$) sólo depende del valor de aquéllas; y 3.º el valor de una cuantificación $\exists x\beta$ o $\forall x\beta$ está determinado por los valores que toma β al asignar a x cada uno de los elementos de A.

Sin embargo, para determinar si una sentencia de segundo orden es verdadera en una estructura **A** debemos salir fuera de **A**; nos hace falta recurrir a la totalidad de subconjuntos de A y a la de relaciones de cualquier número de argumentos entre elementos de A. Y, como vimos en los primeros párrafos de XII para evaluar correctamente las fórmulas de

segundo orden es esencial que consideremos todos estos subconjuntos y todas estas relaciones.

Entendemos qué es un subconjunto de un conjunto y qué una relación entre elementos del conjunto. Pero esto no significa que seamos capaces de precisar qué subconjuntos tiene un conjunto dado y qué relaciones se dan entre sus elementos. Esto es un tema principal de la teoría de conjuntos. De ahí que para resolver problemas en lógica de segundo orden debamos recurrir una y otra vez a las enseñanzas de la teoría de conjuntos (ver Jané, 1988). De hecho, la relación entre lógica de segundo orden y teoría de conjuntos es más íntima de lo que estas frases sugieren (y se mantiene si interpretamos la lógica de segundo orden en términos de cuantificación plural).

Uno de los principios comúnmente admitidos de la teoría de conjuntos es el axioma de elección. En una de sus múltiples formulaciones equivalentes, este axioma dice que toda relación incluye una función con su mismo dominio. Más detalladamente: Si R es una relación y A es un conjunto tal que para todo $a \in A$ hay algún objeto b tal que $R\ (a, b)$ entonces hay una función f tal que para todo $a \in A$, $R\ (a, f(a))$. Así, para cada $a \in A$, f «elige» un objeto, $f(a)$, entre todos aquéllos con los que a se relaciona. El axioma de elección ha sido el más polémico de los axiomas de la teoría de conjuntos (ver Russell, 1919, C. 12) debido a su carácter puramente existencial: afirma la existencia de una función sin decirnos nada acerca de cómo calcularla. Ahora bien, hay una sentencia de segundo orden que es lógicamente válida si y sólo si el axioma de elección es verdadero. Esta sentencia, una formalización del axioma, es

$$\forall X \exists Y(\forall xyz((Yxy \wedge Yxz) \rightarrow y \approx z) \wedge \forall x(\exists y Xxy \rightarrow \exists y(Xxy \wedge Yxy))),$$

donde X e Y son dos variables de predicado binarias. Así, si el axioma de elección es verdadero, la semántica usual de segundo orden le otorga la categoría de verdad lógica.

Algo análogo ocurre con la llamada hipótesis del continuo, según la cual todo conjunto infinito de números reales es o bien numerable o bien tiene la cardinalidad del continuo. La verdad o falsedad de esta hipótesis, propuesta por Cantor, nos es desconocida. Más aún, es independiente de los axiomas usuales de la teoría de conjuntos, y, en la medida en que éstos recogen todo cuanto sabemos acerca de los conjuntos, es independiente de nuestro conocimiento matemático. Sin embargo, hay una sentencia de segundo orden que es lógicamente válida si y sólo si la hipótesis del continuo es verdadera. La sentencia dice: «Para cualesquiera conjuntos (de elementos del universo considerado) X, Y, Z: si X tiene la cardinalidad del continuo, Y es numerable, e $Y \subseteq Z \subseteq X$, entonces Z es biyectable con Y o Z es biyectable con X». No es difícil construir esta sentencia con los medios de V; y no es difícil ver que su validez lógica es equivalente a la verdad de la hipótesis del continuo.

Estos hechos no son excepcionales. Hay muchas proposiciones conjuntistas cuya verdad nos es desconocida y que, sin embargo, es equiva-

lente a la validez lógica de cierta fórmula de segundo orden. La lógica de segundo orden es, pues, de una potencia extraordinaria, ya que tiene la solución a muchos problemas matemáticos. Pero al no poder disponer de un cálculo que nos permita generar las verdades lógicas de segundo orden, esta solución nos es inaccesible.

BIBLIOGRAFÍA

Boolos, G. (1984), «To be is to be a value of a variable (or to be some values of some variables)»: *The Journal of Philosophy, 81*, 430-449.

Boolos, G. y Jeffrey, R. (²1980), *Computability and Logic,* Cambridge University Press, Cambridge-London-New York.

Enderton, H. (1972), *A mathematical introduction to logic,* Academic Press, New York-London.

Gabbay, D. y Guenthner, F. (1983), *Handbook of Philosophical Logic* I, Reidel, Dordrecht-London-Lancaster-Tokyo.

Henkin, L. (1950), «Completeness in the theory of types»: *The Journal of Symbolic Logic, 15*, 81-91.

Jané, I. (1988), «Lógica y Ontología»: *Theoria, 10*, 81-106.

Lewis, D. (1991), *Parts of Classes,* Basil Blackwell, Oxford-Cambridge.

Quine, W. V. (1970), *Philosophy of Logic,* Prentice Hall, Englewood Cliffs.

Russell, B. (1919), *Introduction to Mathematical Philosophy,* George Allen&Unwin, London.

Shapiro, S. (1985), «Second-order languages and mathematical practice»: *The Journal of Symbolic Logic, 50*, 714-742.

Tarski, Mostowski y Robinson (1953), *Undecidable Theories,* North Holland, Amsterdam.

Van Benthem y Doets, K. (1983), «Higher-order Logic», en Gabbay y Guenthner, 1983.

LÓGICA DEÓNTICA

Eugenio Bulygin

I. INTRODUCCIÓN

La lógica deóntica tiene una fecha de nacimiento muy precisa: 1951, año en que aparece el famoso artículo de Georg Henrik von Wright «Deontic Logic» (von Wright, 1951)[1], sin perjuicio de que se pueda encontrar numerosos antecedentes, por lo menos desde el siglo XIV[2]. Entre los antecedentes más recientes cabe mencionar a Bentham, Leibniz y Mally[3].

El impulso inmediato para la creación de una lógica deóntica fue la observación hecha ya por Leibniz y luego, independientemente, por von Wright sobre una serie de analogías sugestivas entre el comportamiento lógico de los conceptos modales aléticos (posible, imposible, necesario) y los conceptos deónticos o normativos de permitido, prohibido y obligatorio. Tomando uno de esos conceptos como primitivo, se puede definir los otros dos de una manera estructuralmente similar. En efecto, si tomamos como primitivo los conceptos de posible (M) y permitido (P), los otros dos pueden ser definidos con la ayuda de la negación. Colocando I por imposible, N por necesario, F por prohibido y O por obligatorio y usando el símbolo habitual para la negación (–), obtenemos el siguiente cuadro:

$I = -M$	$F = -P$
$N = -M-(=I-)$	$Op = -P-(=F-)$

Naturalmente se puede tomar como primitivo cualquiera de los otros dos conceptos, por ejemplo, N (O) y definir los otros en función de éste:

1. En la misma época aparecieron dos trabajos más, dedicados al tema de la lógica deóntica: Becker, 1952 y Kalinowski, 1953. Pero la influencia decisiva se debe a la contribución de von Wright y es por eso que se le considera generalmente, con razón, como el padre de la lógica deóntica.

2. Cf. Knuuttila, 1981. Sobre antecedentes más inmediatos véase Føllesdal-Hilpinen, 1971.

3. Cf. J. Bentham, 1970, Leibniz, *Elementa Juris Naturalis,* 1672, y E. Mally, 1926.

M = – N – P = – O –
I = N – F = O –

El otro aspecto similar está dado por las leyes de distribución: tanto los conceptos modales, como los deónticos se distribuyen de igual manera respecto de la disyunción y la conjunción.

M(pvq)↔MpvMq P(pvq)↔PpvPq
I(pvq)↔Ip&Iq F(pvq)↔Fp&Fq
N(p&q)↔Np&Nq O(p&q)↔Op&Oq

Hay, sin embargo, una diferencia significativa entre el comportamiento lógico de las modalidades aléticas y las correspondientes modalidades deónticas: mientras que en todos los sistemas de lógica modal las fórmulas «p→Mp» y «Np→p» son válidas, las correspondientes fórmulas deónticas «p→Pp» y«Op→p» no pueden serlo. Claramente del hecho de que algo sea (verdad) no se sigue que esté permitido y no todas las obligaciones se cumplen de hecho.

En el primer sistema de von Wright (1951) se acepta que toda tautología de la lógica proposicional es una fórmula válida del sistema cuando las variables proposicionales son reemplazadas por fórmulas deónticas, es decir, fórmulas en las que un operador deóntico es seguido por una expresión de la lógica proposicional. Además, se aceptan las definiciones de prohibido y obligatorio en términos de permisión (que figura como término primitivo), conforme al esquema dado más arriba. Finalmente, se aceptan como axiomas la ley de distribución P(pvq)↔PpvPq y el llamado principio de permisión: PpvP-p (que equivale a Op→Pp). A este sistema lo llamaré el *sistema clásico*. (El mismo von Wright presenta este sistema en forma axiomatizada en von Wright, 1968)[4].

No es mi propósito analizar aquí el desarrollo de la lógica deóntica a partir de la primera obra de von Wright. Dejaré de lado también los problemas de aplicación de la lógica deóntica con sus conocidas paradojas, así como muchos temas que han preocupado a los lógicos deónticos en estos últimos cuarenta años. Me concentraré, en cambio, en el problema de la interpretación de las fórmulas deónticas, que presenta interesantes facetas filosóficas.

II. EL DILEMA DE JØRGENSEN

El tema de la interpretación de la lógica deóntica está en buena medida influenciado por el problema que en los años treinta fue formulado por

4. Por razones de simplicidad paso por alto una complicación: en su presentación originaria von Wright usaba en lugar de variables proposicionales letras mayúsculas (A, B, etc.) interpretadas como nombres de actos y sólo con posterioridad comenzó a usar las variables proposicionales, siguiendo una propuesta de Prior.

el filósofo danés J. Jørgensen y que desde entonces es conocido en la literatura como el *dilema de Jørgensen*[5]. El dilema se apoya en las cuatro tesis siguientes:

1) En el lenguaje corriente se usan en contextos normativos los términos lógicos típicos tales como «no», «y», «o», «si», - «entonces», etc. de la misma manera o al menos de una manera muy similar como en el lenguaje descriptivo, lo que sugiere la idea de considerarlos como conectivas proposicionales. Además, se hacen inferencias en las que las normas figuran como premisas y como conclusiones y tales inferencias tienen todo el aspecto de ser lógicamente válidas. Por lo tanto, hay una lógica de normas que subyace al lenguaje corriente.

2) En la tradición lógica desde Aristóteles hasta nuestros días las relaciones lógicas de implicación (consecuencia lógica) y contradicción se definen en términos de verdad. (Lo mismo ocurre con las conectivas proposicionales.) En consecuencia, sólo expresiones verdaderas o falsas pueden ser objeto del estudio de la lógica.

3) Las normas carecen de valores de verdad.

4) No hay relaciones lógicas entre normas y, por consiguiente, no hay una lógica de normas.

La tesis 4), que se infiere de 2) y 3), contradice la tesis 1), que puede ser considerada como expresión de un hecho preanalítico. Si se quiere evitar la tesis 4) hay que abandonar la tesis 2), o bien la tesis 3). Si, en cambio, se acepta la tesis 4), hay que desarrollar una teoría sustitutiva capaz de reemplazar la lógica de normas para dar cuenta del hecho expresado en 1). Buena parte del desarrollo de la lógica deóntica desde la publicación del primer artículo de von Wright hasta nuestros días puede ser considerado como una discusión del dilema de Jørgensen.

En von Wright (1951), la lógica deóntica es concebida como una lógica de normas y las normas son tratadas —sin ofrecer mayor fundamentación— como entidades verdaderas o falsas. Pocos años después (en el Prefacio a von Wright, 1957) el autor calificó su primer ensayo como «filosóficamente poco satisfactorio» justamente por haber atribuido a las normas valores de verdad y expresó la idea de que la importancia de la lógica deóntica residía precisamente en el hecho de que las normas, aunque alejadas del ámbito de la verdad, están sin embargo sometidas a leyes lógicas («though removed from the realm of truth, yet are subject to logical law»). Esta observación sugiere la ampliación del concepto de lógica y, por lo tanto, el rechazo de la tesis 2), pero von Wright no ha desarrollado luego esta idea[6].

Algunos lógicos y filósofos del derecho han intentado escapar al dilema de Jørgensen mediante el rechazo de la tesis 3), ya sea atribuyendo

5. Cf. Jørgen Jørgensen, «Imperatives and Logic»: *Erkenntnis, 7,* 1937-38.

6. Sólo en un trabajo no publicado, presentado en un simposio de Pisa en 1989, von Wright admite la posibilidad de las relaciones lógicas entre normas, pero le da una fundamentación diferente a la de Alchourrón y Martino, 1990.

valores de verdad a las normas, ya sea postulando valores análogos a la verdad y falsedad, como por ejemplo validez e invalidez (cabe mencionar en este contexto a Kalinowski, Klug y Rödig). Pero ninguno de estos autores ha llevado a cabo una realización satisfactoria de esta idea.

III. INTERPRETACIÓN DESCRIPTIVA DE LA LÓGICA DEÓNTICA

Aquellos autores que en los años cincuenta y en los comienzos de los años sesenta han trabajado en el campo de la lógica deóntica (Prior, Anderson, Lemmon) atribuyeron a las expresiones deónticas valores de verdad, sin preocuparse de la cuestión de si éstas expresaban normas o proposiciones acerca de las normas. Es tan sólo en el transcurso de los años sesenta que esta distinción entre normas y proposiciones normativas fue formulada con claridad (sobre todo en von Wright, 1963), si bien ya se encuentran atisbos de esa distinción en muchas obras anteriores (por ejemplo, en Bentham, Kelsen, Hedenius y Alf Ross). Esa distinción, que parte del hecho de que las oraciones del lenguaje corriente en las que figuran términos típicamente deónticos ('obligatorio', 'prohibido', 'permitido', etc.) son sistemáticamente ambiguas, pues pueden ser interpretados tanto prescriptivamente (como expresiones de normas), como descriptivamente (como expresiones de proposiciones acerca de las normas), abre el camino para la construcción de una lógica deóntica (inobjetable desde el punto de vista de la concepción tradicional de la lógica) como una lógica de las proposiciones normativas. Este camino fue recorrido por von Wright en *Norma y Acción* (1963) al concebir la lógica desarrollada en ese libro como una lógica de las expresiones deónticas interpretadas descriptivamente. Pero su idea fue que la peculiaridad de esa lógica de las proposiciones normativas consiste en que en ella se reflejan las propiedades de las normas mismas. En consecuencia, von Wright propuso en lugar de dos simbolismos diferentes (uno para la lógica de normas y uno para la lógica de las proposiciones normativas) desarrollar un solo simbolismo que admita dos interpretaciones diferentes, una interpretación prescriptiva y otra descriptiva. Esto resultó ser, en mi opinión, un serio error, pues muchos lógicos creyendo que se trata tan sólo de un problema de interpretación han caído en la tentación de interpretar los sistemas clásicos de lógica deóntica como una lógica de las proposiciones normativas. Pero en realidad los operadores deónticos (obligatorio, prohibido, permitido) tienen propiedades lógicas muy diferentes cuando son usados prescriptiva o descriptivamente, es decir, cuando figuran en las normas o en las proposiciones normativas. Por esta razón es imprescindible usar diferentes símbolos. Usaré los símbolos habituales «O» y «P» para los operadores prescriptivos: en consecuencia la fórmula «Op» expresará una norma que ordena p; «O – p» expresará una norma que prohíbe p y «Pp», una norma que permite p. Una orden (esto es, una norma que ordena o hace obligatorio p) exige que p se dé; una prohibición (esto es, una

norma que prohibe p) excluye a p, es decir, exige que p no se dé, y una norma permisiva autoriza que se dé p, es decir, dice que p puede darse.

Los operadores deónticos descriptivos enuncian qué status deóntico tienen determinados estados de cosas o acciones. Este status deóntico lo confieren las normas: cuando una norma N prescribe que p debe ser o se debe hacer (Op), decimos que p es obligatorio en relación a la norma N; cuando una norma N prescribe que p no debe ser o no se debe hacer (O – p), decimos que p está prohibido en relación a la norma N; cuando una norma N prescribe que p puede ser o se puede hacer (Pp), decimos que p está permitido en relación a N.

Podemos también plantear la cuestión bajo qué condiciones una acción o estado de cosas p es obligatorio, permitido o prohibido en relación a un conjunto de normas α. La respuesta es: p es obligatorio en relación a α si y sólo si una norma que prescribe que p debe ser, es decir, una norma de la forma «Op» pertenece a las consecuencias de α. Y p está prohibido en relación a α si y sólo si una norma que prohibe p (es decir, una norma de la forma «O – p») pertenece a las consecuencias de α.

La situación es bastante más complicada en el caso de la permisión. La oración descriptiva «P está permitido en α» es ambigua; a veces lo que se quiere decir con esta oración es que una norma que permite p (es decir, una norma de la forma «Pp») pertenece a las consecuencias de α, pero otras veces la misma oración es usada en un sentido diferente, a saber, en el sentido de que p no está prohibido en α, esto es, que una norma de la forma «O – p» *no* pertenece a α. Esto significa que nos tenemos que ver aquí con dos conceptos de permisión diferentes; los dos operadores permisivos descriptivos serán denominados en lo sucesivo *permisión positiva* y *permisión negativa*[7].

Las proposiciones normativas son siempre relativas a una norma o a un conjunto de normas, esto es, a un sistema normativo; por eso vamos a usar los siguientes símbolos para los operadores deónticos descriptivos: «**O**α», «**P**$^+$α» y «**P**$^-$α», que se leen «p es obligatorio en α», «p está positivamente permitido en α» y «p está negativamente permitido en α», respectivamente.

Las definiciones correspondientes son:

D1. **O**αp = def. «Op»∈Cn(α)
D2. **O**α – p = def. «O – p»∈Cn(α)
D3. **P**$^+$αp = def. «Pp»∈Cn(α)
D4. **P**$^-$αp = def. «O – p»∉Cn(α)

Como ya se ha mencionado, la distinción entre normas y proposiciones normativas ha hecho posible interpretar las expresiones deónticas

7. En trabajos anteriores de Alchourrón y Bulygin fueron usados, siguiendo a von Wright, los términos «permisión fuerte» y «permisión débil» (cf. Alchourrón, 1969 y Alchourrón-Bulygin, 1971 y 1984a). Esa terminología tiene, sin embargo, la desventaja de sugerir la idea de que la permisión fuerte implica a la permisión débil, cosa que en realidad no se da. De ahí el cambio terminológico.

como proposiciones acerca de las normas y con ello construir una lógica deóntica en forma inobjetable desde el punto de vista de la concepción tradicional de la lógica (tesis 2). Esta lógica de las proposiciones normativas fue interpretada como una teoría sustitutiva para la lógica de normas. Pero la caracterización de los operadores deónticos descriptivos dada más arriba muestra claramente que no es suficiente interpretar descriptivamente el simbolismo de los sistemas clásicos de la lógica deóntica; la lógica de las proposiciones normativas exige un simbolismo propio, porque se distingue en aspectos muy esenciales de la lógica de normas.

Las notas distintivas más importantes son las siguientes:

1) Las expresiones de la lógica de las proposiciones normativas son siempre relativas a un sistema, por eso aparecen los suscriptos α. La misma acción p puede naturalmente estar prohibida (permitida, obligatoria) en un sistema normativo y al mismo tiempo no estar prohibida (permitida, obligatoria) en otro. Por eso la proposición normativa «p está prohibido» no es completa: mientras no se indique de qué sistema normativo se trata, esta oración carece de valor de verdad. En cambio, las expresiones de la lógica de normas no están referidas a un sistema normativo; se trata de conceptos absolutos, no relativos[8].

2) En el ámbito del lenguaje prescriptivo no hay nada análogo a la distinción entre la permisión positiva y negativa. Sólo hay un concepto de permisión.

3) Los operadores prescriptivos «O» y «P» son interdefinibles:

$Op = def. -P-p \quad Pp = def. -O-p \quad O-p = def. -Pp \quad P-p = def. -Op$

Contrariamente a la opinión de algunos autores como Weinberger (cfr. C. y O. Weinberger, 1979, 105) esta interdefinibilidad no presupone en modo alguno que el sistema normativo en cuestión esté cerrado y coherente, pues los operadores prescriptivos no están referidos a un sistema determinado, esto es, tienen el mismo significado con independencia del sistema en que figuran.

En cambio, los operadores deónticos descriptivos no son interdefinibles sin más, justamente porque hay dos operadores permisivos distintos. Sólo la permisión negativa es interdefinible con la prohibición: $\mathbf{P}^{-}\alpha p = def. -\mathbf{O}\alpha - p$, pero no la permisión positiva.

4) La definición de los operadores deónticos descriptivos presupone ya la existencia de las relaciones lógicas entre normas: las consecuencias lógicas de α es la clase de todas las normas que se siguen lógicamente de α. Por lo tanto, la lógica de las proposiciones normativas es una extensión de la lógica de normas y los operadores descriptivos se definen en términos de operadores prescriptivos. Sobre este tema volveré más adelante.

5) Finalmente, la negación de los operadores deónticos descriptivos es considerablemente más complicada que la de los prescriptivos. Esto será analizado más detenidamente en la próxima sección.

8. Cf. Carnap, 1942, 41 ss. y 89 ss.

IV. LA NEGACIÓN DE LAS PROPOSICIONES NORMATIVAS

El papel de la negación en el ámbito de las normas es distinto del de la negación en el ámbito de las proposiciones normativas, por de pronto porque hay dos tipos de negación de las proposiciones normativas: la externa y la interna. En el lenguaje corriente la negación de la proposición normativa «p está permitido en α», es decir, «p no está permitido en α» es ambigua: puede significar que α contiene una norma que prohibe p o que α no contiene una norma que permite p. Por lo tanto, resulta conveniente introducir dos signos de negación: «–» para la negación externa y «¬» para la interna, cuyas definiciones son las siguientes:

Negación externa:

$-\mathbf{P}^{+}\alpha p = \text{def. } «Pp» \notin Cn(\alpha)$
$-\mathbf{P}^{-}\alpha p = \text{def. } «-Pp» \in Cn(\alpha)$
$-\mathbf{O}\alpha p = \text{def. } «Op» \notin Cn(\alpha) = \text{def. } «-P-p» \notin Cn(\alpha)$

Negación interna:

$\neg\mathbf{P}^{+}\alpha p = \text{def. } «-Pp» \in Cn(\alpha)$
$\neg\mathbf{P}^{-}\alpha p = \text{def. } «Pp» \notin Cn(\alpha)$
$\neg\mathbf{O}\alpha p = \text{def. } «-Op» \in Cn(\alpha) = \text{def. } «P-p» \in Cn(\alpha)$

De estas definiciones surge que la negación externa de la permisión negativa equivale a la negación interna de la permisión positiva y, en forma similar, la negación interna de la permisión negativa es equivalente a la negación externa de la permisión positiva. Por lo tanto, sólo hay dos formas de negación de la proposición «p está permitido en α» (y no cuatro, como se podría creer), Y también hay dos formas de negación de la proposición «p es obligatorio en α»: la negación externa significa que la norma que ordena p no pertenece a α y la negación interna significa que una norma que permite –p, es decir, una norma de la forma «–Op» (o «P–p») pertenece a α. En otras palabras: la negación externa niega la pertenencia de la norma al sistema, mientras que la negación interna afecta a la norma misma.

Si se considera a la norma «P–p» como norma-negación de «Op» (y correspondientemente, la norma «Pp» como norma-negación de «O–p»)[9], entonces resulta que la negación interna es una operación que lleva de la proposición normativa que afirma la existencia de una norma a la proposición normativa que afirma la existencia de su norma-negación. Pero la negación interna no cumple —a diferencia de la negación externa— los requisitos habituales que se espera debe cumplir una negación. Estos requisitos pueden ser expresados mediante los siguientes cinco postulados[10]:

9. Cf. von Wright, 1983, 133-134.
10. Cf. von Wright, 1963, 138.

1) La negación de una proposición ha de ser una proposición.
2) Tiene que haber una y sólo una negación de una proposición.
3) La negación tiene que ser recíproca, esto es, si una proposición es negación de otra proposición, entonces la segunda proposición ha de ser la negación de la primera.
4) Una proposición y su negación tienen que ser mutuamente excluyentes, es decir, no pueden ser verdaderas las dos.
5) Una proposición y su negación tienen que ser conjuntamente exhaustivas, es decir, no pueden ser falsas las dos.

Claramente sólo la negación externa satisface estos cinco postulados; la negación interna no satisface los postulados 4) y 5), pues tanto la proposición normativa como su negación interna pueden ser ambas verdaderas y también ambas falsas. Son ambas verdaderas cuando el sistema normativo en cuestión es inconsistente (contradictorio) y son ambas falsas cuando el sistema es incompleto. Es justamente la posibilidad de sistemas normativos inconsistentes e incompletos la que hace tan importante la distinción entre operadores deónticos prescriptivos y descriptivos, pues cuando α es un sistema consistente y completo las distinciones entre la permisión positiva y negativa y entre negación externa e interna se desvanecen.

Es interesante investigar el papel de la negación en el ámbito del lenguaje prescriptivo. En los primeros tres postulados es suficiente reemplazar el término «proposición» por el de «norma» para poder aplicarlos a las normas, pues las normas satisfacen claramente estos tres postulados: la negación de una norma es también una norma (por ejemplo, «Op» y «P – p»); para cada norma sólo hay una norma-negación; una norma y su norma-negación son recíprocas (si «Op» es la negación de «P – p», «P – p» es la negación de «Op»).

Los dos últimos postulados, es decir, los postulados 4) y 5) sólo pueden valer para las normas en un sentido analógico, pues las normas carecen de los valores de verdad. Sin embargo, cabe afirmar que una norma como «Pp» («Op») y su norma-negación « – Pp» (« – Op») son mutuamente excluyentes y que las dos normas son conjuntamente exhaustivas, pues las fórmulas «Pp ∨ – Pp» y « – (Pp& – Pp)» son válidas en la lógica de normas, tal como ésta ha sido desarrollada en Alchourrón (1969). Pero es importante darse cuenta cuál es exactamente su significado. Cuando se dice que «Pp» y « – Pp», es decir, una norma permisiva y una norma prohibitiva de p, se excluyen mutuamente, esto no significa que un sistema normativo no pueda contener estas dos normas. Sólo significa que estas dos normas son incompatibles (porque la satisfacción de la prohibición hace imposible hacer uso de la permisión y viceversa, el hacer uso de la permisión de hacer p hace imposible la satisfacción de la prohibición). La lógica de normas establece criterios para la consistencia, pero no dice nada respecto de la existencia de las normas[11].

11. Cf. Alchourrón y Bulygin, 1989, 684-85.

Por razones similares, cuando se afirma que las normas «Pp» y «–Pp» son conjuntamente exhaustivas, esto *no* significa que todo sistema normativo necesariamente contiene una permisión o una prohibición de p. Sólo significa que toda regulación de la acción p implica necesariamente la permisión o la prohibición de p. Sería un error pretender inferir de allí que toda acción esté siempre regulada en todo sistema normativo (y que, por lo tanto, todos los sistemas normativos sean completos y no puedan tener lagunas), justamente porque la lógica de normas nada puede decir sobre los hechos (existencia de normas). En consecuencia, la aceptación de «Ppv–Pp» y «–(Pp&–Pp)» como fórmulas válidas de la lógica de normas no implica en modo alguno que todos los sistemas normativos sean por razones lógicas completos y consistentes, como lo han afirmado algunos filósofos del derecho. Esas fórmulas sólo establecen dos condiciones que las normas han de satisfacer: a) una condición mínima que toda formulación normativa ha de satisfacer para expresar una norma (cuando una formulación normativa no permite ni prohibe la acción p, no expresa ninguna norma respecto de p), y b) una condición para la consistencia, es decir, una condición que toda norma ha de satisfacer para ser consistente (una norma que permite y a la vez prohibe p es contradictoria respecto de p).

He analizado en algún detalle el problema de la negación en la lógica de normas y en la lógica de las proposiciones normativas, porque suele haber no poca confusión respecto de este problema inclusive entre los lógicos que se ocupan de la lógica deóntica[12].

V. LA LÓGICA DE LAS PROPOSICIONES NORMATIVAS

La importancia de la lógica de las proposiciones normativas —que, como surge de las consideraciones anteriores, acusa diferencias importantes respecto de la lógica de las normas— reside en que puede ser entendida como una lógica de los sistemas normativos en el mismo sentido en que la lógica normativa es una lógica de las normas. Las proposiciones normativas son afirmaciones acerca de un sistema normativo —en nuestro simbolismo acerca de Cn(α)— que dicen que determinadas normas pertenecen o no pertenecen a un sistema normativo dado. Una norma pertenece a un sistema normativo cuando o bien ha sido promulgada por alguna autoridad competente del sistema o bien puede ser derivada (es consecuencia lógica) de otras normas que forman parte del sistema. Estas últimas son las normas derivadas. (Para simplificar omito toda referencia a las normas consuetudinarias.)

Las proposiciones normativas se formulan en un lenguaje que es un metalenguaje con respecto al lenguaje en el cual están formuladas las nor-

12. Por ejemplo, en C. y O. Weinberger, 1979 (pp. 121-122) encontramos una noción de negación normativa que no satisface ninguno de los cinco postulados: la negación de una norma no es una norma, la reiteración de la negación no es admisible y la fórmula «Ppv – Pp» no es válida.

mas. Por lo tanto, las oraciones de esta lógica son expresiones metalingüísticas acerca de los sistemas normativos. Es indispensable introducir símbolos especiales para esas expresiones, porque algunas propiedades sumamente importantes de los sistemas normativos, tales como la completitud y la consistencia no pueden ser expresadas adecuadamente en la lógica deóntica tradicional, tanto en su interpretación prescriptiva, como en la descriptiva. En este sentido es interesante comparar los teoremas de los sistemas clásicos (LD) con los de la lógica de las proposiciones normativas (LPN).

Las tesis siguientes son características del sistema DL:

T1. $-(Op \& O-p)$ $(Op \rightarrow -O-p)$
T2. $Pp \leftrightarrow -O-p$
T3. $O(p \& q) \leftrightarrow Op \& Oq$
T4. $P(p v q) \leftrightarrow Pp v Pq$
T5. $Op \rightarrow Pp$
T6. $Pp v P-p$

Estas tesis reflejan las propiedades lógicas de los operadores normativos O y P. Si se los compara con los operadores deónticos descriptivos $\mathbf{O}\alpha$, $\mathbf{P}^{+}\alpha$ y $\mathbf{P}^{-}\alpha$ obtenemos el siguiente cuadro:

1) T1 no es válida en LPN, pues la fórmula «$-(\mathbf{O}\alpha p \& \mathbf{O}\alpha -p)$» puede ser falsa, ya que Cn(α) puede contener tanto «Op», como «O – p» (cuando esto ocurre α es inconsistente).

2) Hay una fórmula análoga a T2, pero sólo para la permisión negativa: $\mathbf{P}^{-}\alpha p \leftrightarrow -\mathbf{O}\alpha -p$.

3) El principio de distribución T3 vale en LPN para $\mathbf{O}\alpha$.

4) El principio de distribución T4 sólo vale para la permisión negativa $\mathbf{P}^{-}\alpha$.

5) T5 vale en cambio sólo para la permisión positiva: «$\mathbf{O}\alpha p \rightarrow \mathbf{P}^{+}\alpha p$» es una fórmula válida. Pero la fórmula «$\mathbf{O}\alpha p \rightarrow \mathbf{P}^{-}\alpha p$» no es válida.

6) La fórmula análoga a T6 no vale, ni para la permisión positiva, ni para la negativa: «$\mathbf{P}^{+}\alpha p$ v $\mathbf{P}^{+}\alpha p$» no es válida en LPN porque el sistema normativo puede tener lagunas, esto es, puede ser incompleto. De 1) se sigue que la fórmula «$\mathbf{P}^{-}\alpha p$ v $\mathbf{P}^{-}\alpha -p$» tampoco es válida, porque el sistema puede ser inconsistente. Esto ocurre cuando α contiene tanto la prohibición de p, como la prohibición de –p, es decir, cuando la acción p es, a la vez, obligatoria y prohibida en el sistema Cn(α). La contradicción consiste en que las dos normas no pueden ser ambas obedecidas por razones lógicas. Cuando una acción es a la vez prohibida y permitida positivamente el sistema normativo correspondiente también es inconsistente. Cabe mostrar que este último caso es un caso especial y no un tipo diferente de inconsistencia (aquí las dos normas son inconsistentes porque es lógicamente imposible hacer uso de la permisión sin violar la norma prohibitiva).

Un estado de cosas p es *normativamente determinado* en un sistema normativo α si, y sólo si, p está o bien permitido positivamente o bien

prohibido en α, esto es, cuando la fórmula «$\mathbf{P}^{+}\alpha p \vee \mathbf{O}\alpha - p$» es verdadera. El concepto de determinación normativa puede servir para la caracterización de los conceptos de laguna y de completitud de los sistemas normativos. Un sistema normativo Cn(α) tiene una laguna o es incompleto cuando un estado de cosas p no está normativamente determinado en α. Sólo cuando todos los estados de cosas (de una cierta clase) están determinados, decimos que Cn(α) es completo (en relación a esa clase).

Puesto que las normas (y muy en especial las normas jurídicas) no sólo pueden ser creadas, sino también anuladas o derogadas, necesitamos un aparato conceptual capaz de dar cuenta del carácter dinámico del orden normativo. La lógica de las proposiciones normativas es adecuada para ello. Un orden jurídico puede concebirse como una secuencia temporal de sistemas normativos que cambian o se modifican con el transcurso del tiempo[13].

VI. EL SISTEMA CLÁSICO COMO LÓGICA DE NORMAS

He tratado de mostrar que no se puede escapar al dilema de Jørgensen recurriendo simplemente a la interpretación descriptiva de las fórmulas de la lógica deóntica. La lógica de las proposiciones normativas tiene sus propias leyes que son muy diferentes de las del sistema estándar. La lógica LPN es una herramienta importante para el análisis lógico de los sistemas normativos, pero no sirve sin más como una teoría sustitutiva de la lógica de normas, en particular, para la justificación de las inferencias normativas. Además ella (al menos en la forma en que ha sido expuesta aquí) presupone ya una lógica de normas, pues sus conceptos fueron definidos en términos de consecuencia lógica y esto implica ya que hay relaciones lógicas entre normas (de lo contrario las normas no tendrían consecuencias lógicas). Por lo tanto, si pensamos que las relaciones lógicas sólo pueden ser definidas en términos de verdad (tesis 2) y que las normas carecen de valores veritativos (tesis 3), estamos de nuevo frente al mismo dilema: o bien abandonamos la tesis 2 o tenemos que desarrollar una teoría sustitutiva para dar cuenta de las relaciones lógicas entre normas. Este último camino fue elegido en Alchourrón-Bulygin (1981) bajo la forma de la concepción expresiva de las normas. Esta teoría tiene su punto de partida en la comprobación de que muchos autores, especialmente los filósofos del derecho, conciben a las normas no como una categoría semántica, sino como una categoría pragmática: lo específicamente normativo estaría dado en el momento pragmático del uso del lenguaje. En Alchourrón-Bulygin, 1981, se intentó investigar esta concepción para determinar su alcance. En esta conceptión las relaciones lógicas no se dan entre las normas (que son actos de ordenar), sino entre sus contenidos, esto es, entre las proposiciones ordenadas. Esto conduce a una lógica

13. Cf. Alchourrón y Bulygin, 1981.

de las proposiciones normativas inobjetable desde el punto de vista de la tradición lógica.

Hoy este intento no me parece totalmente satisfactorio. No tanto porque esta lógica «expresiva» de las proposiciones normativas se apoye en una lógica de normas encubierta, como fue señalado por Weinberger[14], ni tampoco porque la concepción expresiva no haya podido hasta ahora ofrecer una teoría satisfactoria de las normas condicionales, sino básicamente porque la justificación de una sentencia judicial —que tiene carácter normativo— requiere premisas normativas. Esto significa que el juez ha de derivar su decisión de las normas mismas y no de meras proposiciones acerca de las normas. Por eso una lógica de normas es imprescindible.

Qué aspecto ha de tener una genuina lógica de normas ya fue señalado en Alchourrón, 1969. Más tarde, estos análisis fueron usados en Alchourrón-Bulygin, 1971, donde fueron construidos dos sistemas lógicos, uno para la lógica de normas y otro para la de las proposiciones normativas. Esto puso de manifiesto un hecho interesante, a saber, que el sistema estándar de lógica deóntica resultó ser una reconstrucción básicamente correcta de las propiedades lógicas de los operadores normativos O y P. Pero es fundamental que esta lógica de normas sea suplementada con una lógica de las proposiciones normativas.

Sin embargo, ni en Alchourrón, 1969, ni en Alchourrón-Bulygin, 1971, se encuentra una fundamentación satisfactoria de la lógica de normas. Se trabaja allí con conectivas proposicionales y se habla de relaciones lógicas entre normas, sin explicar de qué manera esto es compatible con el hecho de que las normas carecen de valores de verdad. Y es significativo que en los últimos años justamente von Wright, uno de los fundadores de la lógica deóntica, se ha vuelto escéptico respecto de su posibilidad.

Si se acepta que las normas carecen de valores de verdad, no cabe duda de que una lógica de normas genuina sólo es posible si se amplía el concepto de lógica de tal manera que las conectivas proposicionales y los conceptos de implicación (consecuencia) lógica y de consistencia puedan ser definidos sin hacer referencia a la noción de verdad. Una propuesta en tal sentido fue formulada recientemente en Alchourrón-Martino, 1990. Estos autores proponen definir la noción de consecuencia lógica sobre la base del concepto abstracto de consecuencia (caracterizado por Tarski), que se usa como concepto primitivo y que no es ni sintáctico, ni semántico. Las conectivas proposicionales se definen luego a la manera de Gentzen mediante reglas de introducción y eliminación. Para eludir los peligros señalados en Prior, 1960, tales reglas se introducen en el sentido de Belnap, 1962, en un contexto de deducción caracterizado axiomáticamente.

Esta propuesta consiste fundamentalmente en justificar la idea, ya expresada en von Wright, 1957, de que el campo de la lógica es más

14. Cf. Weinberger, 1984a.

amplio que el de la verdad. Ésta no será objeto de análisis en este trabajo, pero es claro que si tal propuesta resultara viable, se lograría un terreno firme para fundamentar una auténtica lógica de normas.

BIBLIOGRAFÍA

Alchourrón, C. E. (1969), «Logic of norms and logic of normative propositions»: *Logique et Analyse, 12,* 242-268.

Alchourrón, C. E. y Bulygin, E. (1971), *Normative Systeme,* Springer Verlag, Wien-New York.

Alchourrón, C. E. y Bulygin, E. (1979), *Sobre la existencia de las normas jurídicas,* Universidad de Carabobo, Valencia (Venezuela).

Alchourrón, C. E. y Bulygin, E. (1981), «The expressive conception of norms», en R. Hilpinen (ed.), *New Studies in Deontic Logic,* Reidel, Dordrecht-Boston-London.

Alchourrón, C. E. y Bulygin, E. (1984a), «Permission and permissive norms», en W. Krawietz *et al.* (eds.), *Theorie der Normen,* Duncker & Humblot, Berlin.

Alchourrón, C. E. y Bulygin, E. (1984b), «Pragmatic foundations for a logic of norms»: *Rechtstheorie, 15,* 453-464.

Alchourrón, C. E. y Bulygin, E. (1989), «Von Wright on Deontic Logic and the Philosophy of Law», en P. A. Schilpp and L. E. Hahn (eds.), *The Philosophy of Georg Henrik von Wright,* The Library of Living Philosophers, La Salle.

Alchourrón, C. E. y Martino, A. A. (1990), «Logic without Truth»: *Ratio Juris, 3,* 46-67.

Anderson, A. R. (1956), «The Formal Analysis of Normative Systems», en N. Rescher (ed.), *Logic of Decision and Action,* Pittsburgh, 147-213.

Becker, O. (1952), *Untersuchungen über das Modalkalkül,* Meisenheim am Glan.

Belnap, N. D. (1962), «Tonk, Plonk and Plink»: *Analysis,* 130-134.

Bentham, J. (1970), *Of Laws in General* (ed. H. L. A. Hart), The Athlone Press, London.

Carnap, R. (1942), *Introduction to Semantics,* Harvard University Press, Cambridge.

Føllesdal, D. y Hilpinen, R. (1971), «Deontic Logic: An Introduction», en R. Hilpinen, *Deontic Logic: Introductory and Systematic Readings,* Reidel, Dordrecht.

Kalinowski, G. (1953), «Théorie des propositions normatives»: *Studia Logica, 1,* 147-182.

Knuuttila, S. (1981), «The Emergence of Deontic Logic in the Fourteenth Century», en R. Hilpinen (ed.), *New Studies in Deontic Logic,* D. Reidel, Dordrecht-Boston-London.

Lemmon, E. J. (1965), «Deontic Logic and the Logic of Imperatives»: *Logique et Analyse, 8,* 39-71.

Mally, E. (1926), *Grundgesetze des Sollens: Elemente der Logik des Willens,* Graz.

Prior, A. N. (1960), «The runabout inference-ticket»: *Analysis, 21,* 38-39.

Prior, A. N. (1962), *Formal Logic,* Clarendon Press, Oxford.

Weinberger, C. y O. (1979), *Logik, Semantik, Hermeneutik,* C. H. Beck, München.

Weinberger, O. (1984), «On the Meaning of Norm Sentences, Normative Inconsistency and Normative Entailment. A Reply to Carlos E. Alchourrón and Eugenio Bulygin»: *Rechtstheorie, 15,* 465-475.

Wright, G. H. von (1951) «Deontic Logic»: *Mind, 60,* 1-15.

Wright, G. H. von (1957), *Logical Studies,* Routledge and Kegan Paul, London.

Wright, G. H. von (1963), *Norm and Action,* Routledge and Kegan Paul, London.

Wright, G. H. von (1968), *An Essay in Deontic Logic and the General Theory of Action,* Acta Philosophica Fennica XXI, Amsterdam.

Wright, G. H. von (1983), «Norms, Truth and Logic», en G. H. von Wright, *Practical Reason,* Oxford, 130-209.

LÓGICA E INTELIGENCIA ARTIFICIAL

Raúl J. Carnota

> ... there is very little difference between traditional work in philosophical logic and logical research in theoretical AI...(R. Thomason).

I. INTRODUCCIÓN

En las primeras épocas de la Inteligencia Artificial (IA), dos de sus pioneros señalaban:

> ...A computer program capable of acting intelligently in the world must have a general representation of the world in terms of which its inputs are interpreted. Designing such a program requires commitments about what knowledge is and how it is obtained. Thus, some of the major traditional problems of philosophy arise in artificial intelligence... More specifically we want a computer program that decides what to do by inferring in a formal language that a certain strategy will achieve its assigned goal. This requires formalizing concepts of causality, ability, and knowledge. Such formalisms are also considered in philosophical logic... (McCarthy y Hayes, 1969).

Esta cita es representativa de la fascinación que ejerció la Lógica sobre una buena parte de los pioneros de la IA. Sin embargo, durante mucho tiempo existió un cierto aislamiento entre los dos campos, hasta que a fines de los 70 y comienzos de los 80, el auge de la IA promovió el interés de científicos de disciplinas como la filosofía, la lógica, la lingüística y la psicología, que se introdujeron, con su bagaje previo, en el nuevo campo.

A partir de ese momento, las interconexiones comenzaron a ser más ricas: no sólo desde la IA se utilizaban los resultados de la Lógica, sino que las necesidades surgidas de la aspiración de inculcar raciocinio «inte-

ligente» a una máquina y la intención de hacerlo a través de reconstruir formalmente los procesos del razonamiento, motivaron a muchas personas relevantes del área de la lógica.

Este artículo intenta reflejar algunas de las facetas de estas relaciones, poniendo el énfasis en cómo las dificultades halladas en los intentos de uso de la Lógica Clásica, reverdecen discusiones que vienen de antiguo en el campo filosófico y en cómo el objetivo de producir el diseño de sistemas computacionalmente operables provoca al interior de la Lógica un nuevo estímulo para responder a viejas preguntas en una forma que debe ahora ser mucho más concreta y no discursiva.

No pretendemos ignorar que la IA está atravesada por distintas corrientes, cada una de ellas con sus puntos de vista notoriamente divergentes acerca de cómo llevar adelante los objetivos del campo, y que muchos de sus más prominentes investigadores no son, ni mucho menos, partidarios del uso de la Lógica (incluso en sus variantes no clásicas) para reconstruir la «inteligencia» en los robots.

Las propias expresiones de McCarthy-Hayes que citamos antes, podrían ser refutadas por otros investigadores y la referencia al «conocimiento» que el programa posee como el resultado de un proceso inferencial en un lenguaje formal sería señalada, más como una limitación (en el sentido de sus hipótesis acerca de la inteligencia humana), que como una aspiración.

No abordaremos aquí esta polémica (salvo una mínima referencia en la sección II) y consideraremos sólo los enfoques «proposicionales» de representación y manipulación de la información que han sido los estimuladores del desarrollo de nuevos formalismos basados en lógica. Una discusión extensa de los fundamentos de la IA y de las distintas corrientes que coexisten en la disciplina puede encontrarse, entre otras fuentes, en números especiales de las revistas *Dedalus* (1987) y *Artificial Intelligence, AI* (1991).

Un objetivo esencial en la IA es la formalización del «razonamiento de sentido común». En la vida cotidiana extraemos conclusiones en base a generalizaciones que tienen excepciones («los pájaros vuelan», «si el auto tiene combustible y batería, arranca»). Sin embargo estas conclusiones, ante la adquisición de nueva información (que el pájaro en cuestión es un pingüino, que existe un cable cortado en el motor), pueden resultar erróneas.

La representación directa en Lógica Clásica (LC) de estos razonamientos trae problemas. Si tenemos:

$\forall x$ (Mamífero(x) $\rightarrow$ — Vuela(x)), y
$\forall x$ (Murciélago(x) $\rightarrow$ Vuela(x)),

se deriva en LC que no pueden existir individuos que sean a la vez mamíferos y murciélagos. Esto es debido a que la LC es monótona en el sentido de que, al ser las premisas condición suficiente para la conclusión, el agregado de nuevas premisas no puede invalidar nunca la conclusión.

Los intentos de formalización del Razonamiento No Monótono (RNM) son, quizás, el caso más claro de desarrollos que han afectado al propio campo de la lógica, tal vez por sus conexiones con viejos problemas ya planteados en aplicaciones científicas (suposiciones «ceteris paribus»), deónticas (obligaciones «prima facie»), del discurso cotidiano (condicionales contrafácticos), etc. Su análisis será el tema central del presente trabajo.

La búsqueda de inspiración para el diseño de sistemas inteligentes en las características de la inteligencia humana es una fuente de otras críticas y de desafíos que afectan a la lógica. La constatación de que los seres humanos no siempre somos consistentes ha motivado la aplicación de lógicas paraconsistentes. El convencimiento de que la mente humana toma decisiones sin analizar todas las alternativas (sorteando los problemas de complejidad computacional por medio de supuestos) es también una incitación al desarrollo de procedimientos y lenguajes que reconstruyan esa *performance*. Lo mismo puede decirse de las caracterizaciones vagas o difusas, los razonamientos por analogía, el aprendizaje como proceso inductivo, etc. Si bien es imposible en pocas páginas desarrollar todos estos aspectos, vale su mención por tratarse de cuestiones vivas y actuales en la investigación conjunta de Lógica e Inteligencia Artificial. Finalmente, la Lógica es usada, también, como «metalenguaje» de los lenguajes de representación, esto es, como herramienta para caracterizar el poder inferencial de un lenguaje, tal como ha sido propuesto por A. Newell, en Newell (1981), al caracterizar el «nivel del conocimiento» en los sistemas inteligentes.

En la próxima sección se hace un breve repaso de los objetivos de la IA. La sección III presenta, bajo la hipótesis de una representación «declarativa» del conocimiento, las conexiones «naturales» entre la Lógica y la IA. La sección IV reseña los principales cuestionamientos al uso de la Lógica Clásica, en particular a su capacidad para capturar el razonamiento «de sentido común». En la sección V, la central del trabajo, se discuten algunas cuestiones lógicas ligadas a la formalización del raciocinio no monótono. Finalmente, en la sección VI se realizan algunas reflexiones finales sobre el rol de la Lógica en la IA.

II. INTELIGENCIA ARTIFICIAL

> ...Most practitioners would agree on two main goals in AI. The primary goal is to build an intelligent machine. The second goal is to find out about the nature of intelligence... (R. Schank).

Desde tiempos inmemoriales los seres humanos han imaginado artefactos animados —ídolos, imprevisibles dioses, obedientes esclavos, robots—,

que compartieran la «esencia humana». Qué componía esa «esencia» es algo que fue cambiando con el tiempo, pero con un elemento invariante: ser humano es pensar, razonar, asociar, crear.

Mitos, historias, argumentos filosóficos o búsquedas científicas —como elixir de la vida— reflejan a través de los siglos esa obsesión.

Pero recién con la existencia de las computadoras, los primeros instrumentos aptos para procesar símbolos en forma totalmente general, es que esta búsqueda mítica pasa a constituirse en una zona de investigación formal, con posibilidad de definir rigurosamente aspectos de la actividad inteligente, testearlos y lograr la realimentación rápida que permita el avance de la experimentación. Y hemos dicho aspectos, porque la definición de inteligencia es una bruma que envuelve a cualquiera que desee comenzar a caracterizar el área.

Desde su fundación formal en la Conferencia de Darmouth en 1956, la investigación en IA se ha realizado siguiendo dos enfoques conectados y enfrentados mutuamente. El primero tiene por meta principal la construcción de sistemas orientados a la resolución de problemas, sin necesariamente imitar la forma en que la mente humana realiza esta tarea aunque sí buscando alcanzar su *performance.* El segundo intenta imitar los modos de funcionamiento de la inteligencia humana por medio de un programa de computación, buscando arrojar luz sobre el proceso cognitivo humano

En un caso se busca producir conductas que podrían ser producidas a través del uso de la inteligencia, independientemente de los medios empleados en obtener el resultado. Si la conducta obtenida tiene un grado considerable de inteligencia, la simulación es exitosa.

En el otro se pretende construir modelos. Un modelo debe producir también una salida apropiada, pero debe hacerlo mediante procesos y representaciones de información que sean espejo de los procesos inteligentes que se producen en la mente humana.

En esta dirección la corriente que iniciaron en los 40 McCulloch y Pitts y que se inspiraba en los paralelos entre la naturaleza binaria de las neuronas y los componentes electrónicos de las computadoras, luego de casi dos décadas de marginalidad, revive en la investigación actual en redes neuronales. Sin embargo, más que por aquella analogía, mucho del impulso actual de esta corriente se explica por su aparente capacidad para encarar eficientemente problemas que el enfoque simbólico no ha resuelto.

1. *El paradigma simbólico*

> A physical symbol system has the necessary and sufficient means for general intelligent action. By necessary, we mean that any system that exhibits general intelligence will prove upon analysis to be a physical symbol system. By sufficient we mean that any physical symbol system of sufficient size can be organized further to exhibit general intelligence (Newell y Simon, 1976).

En la primera corriente, Newell y Simon suponen que el cerebro humano y la computadora, siendo totalmente distintos en estructura y en mecanismos, tienen una descripción funcional común en cierto nivel de abstracción. En ese nivel ambos pueden ser vistos como ejemplos de un «aparato» que genera conductas inteligentes manipulando símbolos por medio de reglas formales.

En la cita previa, *physical* significa que dichos sistemas obedecen las leyes de la física —en particular son construcciones ingenieriles concretas—. Por *symbol system* se entiende una colección de patrones y procesos. Los procesos son capaces de producir, destruir y modificar los símbolos. Los patrones tienen la capacidad de designar objetos, procesos u otros patrones. Cuando un patrón designa a un proceso, puede ser «interpretado», lo que implica llevar adelante el susodicho proceso.

Inteligencia es, en este punto de vista, la habilidad para procesar símbolos.

Dentro del paradigma simbólico, se pueden distinguir numerosas posturas diferentes sobre la naturaleza de los procesos mentales y las subsecuentes actividades de los científicos de la IA. En un extremo se encuentran los llamados «logicistas», cuyas posiciones están desarrolladas en Nilsson (1991), que sostienen la posición de que la «inteligencia» está basada en una representación declarativa de las creencias acerca del mundo y que la actividad cognitiva está basada esencialmente en procesos inferenciales. En este enfoque, la cuestión determinante para un diseñador de una Base de Conocimientos sería la conceptualización del «mundo» que quiere reflejar. Esta conceptualización se volcaría luego en esta Base bajo la forma de un conjunto de sentencias de algún lenguaje proposicional. La representación así alcanzada aspira a ser lo más independiente posible del modo en que puede llegar a ser usada la información.

Alrededor de la aceptación o rechazo de algunas de estas hipótesis giran otras corrientes de investigadores. Polemizando con Nilsson, en Birnbaum (1991) se resalta la imposibilidad de una caracterización del conocimiento totalmente independiente del uso y se propone, como alternativa a la semántica de la teoría de modelos, una «semántica funcional», basada en la idea de que la representación toma sentido en función de su rol causal en los procesos mentales, y, en última instancia, en la percepción y en la acción. El concepto de número primo, ejemplifica, no significa lo mismo para mí (Birnbaum), que para un matemático especializado en teoría de números. Y su sentido cambiará también para mí si comienzo a estudiar alguna de tales teorías. Con esta aproximación, no se puede lograr una especificación de lo que un organismo conoce independientemente de lo que hace.

En el resto de este trabajo vamos a adoptar un enfoque de tipo «simbolista», por ser el marco en el cual se producen las conexiones más ricas entre la Lógica y la IA.

En particular, al hablar de estructuras de Representación del Conocimiento (RC) en un sistema, nos limitaremos a las «proposicionales»,

es decir, estructuras de datos interpretables como fórmulas lógicas de algún tipo.

De acuerdo con Levesque (1986a), este tipo de estructuras deben ser interpretables proposicionalmente, es decir, como expresiones en un lenguaje con una teoría de verdad. Debe ser posible señalar una de dichas estructuras y decir cómo debe ser el mundo para que ella sea verdadera.

2. *La Inteligencia Artificial como interdisciplina*

En todo caso, más allá de las facciones, la IA se constituye crecientemente en una zona de convergencia de las más diversas disciplinas: la lógica, la psicología, la filosofía, la economía, la matemática y, también, las ciencias de la computación. Un resultado valedero en IA lo es en la medida en que aporte en alguno de los aspectos que requerimos para considerar un agente como inteligente: capacidad de comunicación, conocimiento de sí mismo (conocimiento de su conocimiento), conocimiento del mundo, acumulación de experiencia para reusarla en la interacción con él, intencionalidad en sus objetivos (capacidad de construir y adaptar planes según las circunstancias), creatividad (al menos en algún sentido débil, como la adaptación a cambios en el ambiente, en suma, aprendizaje). Tópicos que, sin duda, están mejor tratados en libros de psicología o filosofía que en los textos de IA. En consecuencia, si ese resultado existe, deberá ser reconocido «fuera» de la IA, como un aporte. La característica específica de estos aportes de la IA es la mecanización, particularmente cuando las formas standard de mecanización han demostrado ser intratables (en el sentido de la complejidad computacional).

Todos los campos del conocimiento son, en alguna medida, IA. Todos tratan acerca de la naturaleza del hombre. La importancia de la IA está dada en la medida en que sus aportes tecnológicos sean significativos. Las preguntas que intentamos responder es lo único que realmente importa.

III. LA CONEXIÓN CON LA LÓGICA

> ...logic is at the heart of reasoning, and reasoning is at the heart of intelligence (W. J. Rapaport).

Como ya vimos, en el contexto del paradigma simbólico de la IA, un sistema «inteligente» es aquel con habilidad para procesar símbolos mediante reglas formales.

En un sistema que se inspire en las funciones inteligentes de los humanos, podemos distinguir varios tipos de procesamientos de símbolos. Las funciones perceptuales, que detectan los distintos datos del mundo; los procesos de memorización, que almacenan y organizan la información

y luego son capaces de recuperarla de acuerdo a ciertos objetivos; los que podríamos llamar procesos deliberativos, el núcleo del «razonamiento», en los que se construyen hipótesis, se analizan alternativas, se deciden caminos, etc.; y las funciones efectoras que actúan sobre el mundo de acuerdo a las acciones decididas en los procesos deliberativos.

Para construir un sistema computacional que realice estas actividades, se precisan definir ciertos procedimientos destinados a cumplir esas funciones. En particular nos interesan los procedimientos destinados a elaborar la información.

Aceptemos que las unidades mínimas en que se recoge y almacena la información son los enunciados de algún lenguaje formal. El asunto no es trivial, ya que según como sea el instrumento que usemos para receptar y almacenar la información, será el modo en que podamos elaborarla.

Supongamos que $A_1 \ldots A_n$ son distintas oraciones del lenguaje elegido, almacenadas mediante algún procedimiento. Ese conjunto finito de oraciones representará cierta información significativa que interesa retener.

Convengamos que «elaborar inteligentemente» esa información significa obtener nueva información a partir de la misma. Esta nueva información será representada por enunciados, en principio diferentes de los de partida y que, de alguna manera, están vinculados a éstos. Estos enunciados amplían la información que el sistema posee.

Si $A_1,\ldots,A_n$ es la información de partida, llamaremos B a la información que se alcanza luego de un número finito de pasos.

Esta imagen secuenciada de pasos en los que la información poseída se va enriqueciendo, tiene una cierta connotación psicológica y no es extraño que haya sido tomada como una metáfora del funcionamiento del raciocinio humano.

¿De qué manera, bajo qué condiciones y qué cosas se agregan en cada paso a la Base de Conocimientos existente? Bajo nuestras hipótesis de partida, esto tiene que estar bien especificado por medio de «reglas formales», para que la máquina sepa qué hacer frente a cada situación o estado. Cada información agregada se vincula a algunas de las informaciones existentes (eventualmente a todas). Por ejemplo, cierto B_1 se agregará a resultas de la presencia de A_1 y A_2. Llamemos al procedimiento que permitió este paso Regla 1 (regla de transformación 1). Así, cada uno de los pasos estarán justificados por algún conjunto R_1, $R_2,\ldots$, R_m de reglas de transformación, cuyo sentido es incorporar nuevas informaciones a partir de las originales.

El proceso así descrito no es otra cosa que lo que la lógica siempre tuvo en su historia como su objetivo: cómo (bajo qué condiciones) justificar ciertos enunciados, apoyándose en otros enunciados. Siguiendo la tradición escolástica, dadas las oraciones:

Sócrates es un hombre (A_1) y
Todos los hombres son mortales, (A_2)

cómo justificar la oración

Sócrates es mortal (B)

Si tenemos los enunciados $A_1,\ldots,A_n$ y se ha podido construir un proceso de transformaciones que terminan en B, anotamos:

(*) $A_1,\ldots,A_n \vdash B$,
que se lee «B es conclusión sintáctica de las premisas $A_1,\ldots,A_n$».

Lo que (*) representa es abreviatura de la siguiente afirmación existencial: existe una secuencia de enunciados del lenguaje utilizado, tal que el último de la secuencia es B, y tal que los enunciados de la secuencia, o bien son los A_i, o bien se van construyendo, a partir de los anteriores, en base a reglas de transformación (reglas de inferencia en el lenguaje de la lógica).

La clave de lo enunciado hasta aquí, parece estar en las reglas de transformación.

Si lo único que le pedimos a una regla es que transforme unos enunciados en otros, una regla podría, por ejemplo, convertir enunciados que comienzan en vocal en otros que terminan con consonante.

Ver a las reglas lógicas como reglas de transformación es correcto. Pero no cualquier regla de transformación es considerada una regla lógica.

¿Cuál es el «control de calidad» que tenemos que reclamar de las reglas de transformación?

La respuesta de la filosofía de la lógica contemporánea está en el concepto de consecuencia lógica. Para evitar meras transformaciones sintácticas que puedan llevar a verdaderos absurdos, las transformaciones deben tener cierta calidad que, según Tarski (1956), puede formularse así:

(1) Si $A_1,\ldots,A_n \vdash B$ entonces $A_1,\ldots,A_n \models B$ y se lee «para que B sea una conclusión de A_i "aceptada", debe ocurrir que B sea lógicamente implicado por los enunciados A_i».

La noción de «lógicamente implicado» se define así: B está lógicamente implicado por los A_i si, en cualquier contingencia posible en la que los A_i sean verdaderos, B no tenga más remedio que ser verdadero. O sea que la verdad se hereda a todo lo que se va obteniendo como resultado de la aplicación de las reglas de transformación. O también que B tenga justificada su verdad toda vez que se tenga justificada la verdad de los A_i. Esta relación de justificación la tendrá cada enunciado de la secuencia respecto a los enunciados de partida y es una condición adicional en la construcción de la secuencia.

La idea central de la lógica vira hacia el estudio de estas «condiciones de calidad» de las inferencias. Una inferencia será «buena» si cumple con la condición de (1). Se está pidiendo que la conclusión tenga garantizada su verdad a partir de la verdad de las premisas.

La cuestión conversa es pedir que, en el caso de que la verdad de B esté garantizada por la verdad de los A_i, se pueda construir una secuencia, mediante reglas de transformación, que empiece por los A_i y termine con B:

(2) Si $A_1,\ldots,A_n \models B$ entonces $A_1,\ldots,A_n \vdash B$.

Cómo encontrar, dado B, y un conjunto de reglas de transformación, una secuencia de pasos que partiendo de nuestras «premisas», llegue a B, es el problema de la decisión para la noción de consecuencia sintáctica.

De lo expuesto hasta aquí se podría concluir que existe una casi total superposición entre las actividades de la lógica y los procedimientos de elaboración de la información en un sistema de IA.

Aquí surge, sin embargo, una pregunta:

¿Son las restricciones expresadas en (1) y (2) adecuadas respecto a los objetivos de la IA? Éste es el tema que discutiremos en la próxima sección.

IV. LOS CUESTIONAMIENTOS A LA LÓGICA DEDUCTIVA

> ...logical reasoning is more appropiate for displaying or confirming the results of thinking than for thinking itself... (M. Minsky).

¿Cómo impartir nociones de sentido común a un robot?

¿Cómo diseñar un robot con una capacidad de razonamiento suficientemente poderosa y útil como para que, una vez provisto de un subconjunto de ese «sentido común», sea capaz de generar suficiente del resto como para adaptarse a su entorno y operar inteligentemente sobre él?

¿Será la Lógica el marco adecuado para afrontar este desafío?

Partiendo de su referente básico, la naturaleza del raciocinio humano, muchos investigadores de la IA encuentran a la lógica como demasiado formal y limitada, y perciben que los procesos de razonamiento abarcan un espectro mucho más amplio que el análisis lógico deductivo.

Curiosamente, estos cuestionamientos no son nuevos:

Stuart Mill (citado en Cohen y Nagel, 1957), afirmaba que era incontestable que, en el ejemplo de Sócrates, (B) está presupuesta en (A_1) y que —peor aún— no podemos asegurar la mortalidad de todos los hombres, a menos que estemos convencidos de la de cada hombre en particular.

Por su parte, en Cohen y Nagel (1957), se formula así la llamada «paradoja de la inferencia»: Si en una inferencia la conclusión no está contenida en la premisa, la inferencia no puede ser válida; y si la conclusión no es diferente de las premisas, la inferencia es inútil; pero la conclusión no puede estar contenida en las premisas y al mismo tiempo poseer novedad; en consecuencia las inferencias no pueden ser a la vez válidas y útiles.

M. Minsky, exponente de los que consideran inadecuada a la postura logicista (entendida como la reducción del proceso de construcción de un sistema inteligente a la selección de un conjunto de sentencias en un lenguaje lógico, sobre el que opera un demostrador de teoremas como emulador de los mecanismos de razonamiento), sintetiza, en Minsky (1975), sus cuestionamientos.

Los principales problemas que señala son los siguientes:

1. El problema de la relevancia. La información contenida en una Base de Conocimientos (BC) precisa de meta-información que indique en qué circunstancias y de qué manera va a ser usada. Asimismo un sistema inteligente debería tener claro cuándo es pertinente o no realizar ciertas deducciones.

2. El problema de la monotonía, que ya se ilustró en la Introducción.

3. La separación entre axiomas y deducción (que equivale a independizar el conocimiento de su forma de uso), dificulta la clasificación de las proposiciones y el control del proceso deductivo. Si, por ejemplo, se desea axiomatizar la relación de proximidad, resulta natural que sea transitiva:

$$(A \text{ prox } B) \wedge (B \text{ prox } C) \rightarrow (A \text{ prox } C).$$

Sin embargo, una aplicación irrestricta de esta regla, puede conducir a que todo esté próximo a todo. En un sistema lógico «puro» no es fácil hacer un nuevo axioma que «prohíba» aplicar la transitividad más de cierto número de veces.

4. La explosión combinatoria. Los sistemas basados en lógica no escapan al problema de la explosión combinatoria, cuando se encaran dominios complejos.

5. La exigencia de consistencia. En un sistema basado en la LC, no es posible representar una sentencia y su negación, sin que el proceso inferencial «trivialice» las conclusiones (permita derivar cualquier sentencia). Esta propiedad, lejos de ser una virtud, es, para M. Minsky, ni necesaria ni deseable, ya que hace que los sistemas así constreñidos resulten muy débiles en relación al poder de raciocinio de los agentes humanos inteligentes. Nadie es completamente consistente. Lo que es importante es cómo el agente maneja paradojas o conflictos, cómo aprende de los errores, cómo intenta sortear las situaciones de las que sospecha que puedan resultar inconsistentes.

Minsky reconoce la necesidad del uso de mecanismos deductivos, pero los circunscribe:

> ...I do not mean to suggest that «thinking» can be proceeded very far without something like «reasoning». We certainly need (and use) something like syllogistic deduction; but I expect mechanisms for doing such things to emerge in any case from processes for «matching» and «instantiation» required for other functions. Traditional formal logic is a technical tool for discussing either everything that can be deduced from some data or whether a certain consequence can be so deduced; it

cannot discuss at all what ought to be deduced under ordinary circunstances. Like the abstract theory of Sintax, formal Logic without a powerful procedural semantics cannot deal with meaningful situations... (Minsky, 1975).

Como se ve, Minsky señala las dificultades de la LC, tanto para la representación de propiedades «típicas», como para decidir qué debería inferirse en condiciones «típicas» o «normales».

1. *La relevancia*

> ...The logical view of thinking has considerable attraction, since logic offers valuable insight into the relation between rational held beliefs. Unfortunately, it does not cover all such relations. Its more striking omission is any consideration of the purpose (or purposes) of reasoning. That is, the logical view of thinking ignores questions of whether one should or should not draw some inference, and whether one inference is better or more appropiate than another... (J. Doyle).

La lógica propone inferencias seguras, pero no siempre las útiles para determinados propósitos. Una inferencia apropiada en un dominio, puede ser irrelevante en otro. Similares reclamos se realizan desde el campo filosófico (teoría de la argumentación). En ambos casos se postula la necesidad de una «racionalidad útil», ya que el mero razonar correcto —que es lo que garantizan los procesos deductivos— puede ser «irracional» considerando los propósitos del razonador y el dominio en que está operando.

Un concepto de «racionalidad útil», presentado en Doyle (1989), postula que cada paso de razonamiento sea dado en el sentido de maximizar la función de valor que el agente establece en base a sus expectativas y preferencias. En ese caso, antes que formular la pregunta acerca de si las inferencias son «seguras», importa preguntarse si los pasos de razonamiento dados y las conclusiones alcanzadas sirven eficientemente a los propósitos del razonador.

En otras palabras, si el razonar del robot tiene como paradigma el razonar humano y quiere, al menos, imitar su *performance* no le servirá cualquier inferencia, por más que sea lógicamente válida, si no es relevante para sus objetivos.

Contrario sensu, hay situaciones en las que el sistema debe actuar aunque no posea una descripción completa del estado de cosas existente o aunque, en caso de poseerla, no resulte tolerable en el tiempo el análisis de todos los factores en juego. En este caso actuará a partir de extraer conclusiones no seguras (en el sentido del «control de calidad» exigido por la lógica standard), corriendo el riesgo de que, ante la llegada de nueva información, las decisiones adoptadas hayan resultado erróneas. Esta contracara de la racionalidad útil nos lleva a la cuestión de las extensiones no monótonas de las inferencias deductivas.

2. *La monotonía*

> ...A key property of intelligence —whether exhibited by man or by machine— is flexibility. This flexibility is intimetely connected with the defeasible nature of commonsense inference(s)...; we are all capable of drawing conclusions, acting on them, and then retracting them, if necessary in the face of new evidence. If our computer programs are to act intelligently, they will need to be similarly flexible... (M. Ginsber).

Supongamos que un amigo nos menciona un pájaro, llamado Pi-pio, que tiene en su casa. Seguramente lo imaginaremos en una jaula ya que, si no tenemos más información sobre el mismo, es razonable suponer que vuela.

¿Cómo representar esta asociación heurística?

Un camino en Lógica de primer orden es la formulación «Todos los pájaros vuelan»:

$\forall x$ Pájaro(x) $\rightarrow$ Vuela(x) (1).

¿Qué ocurre en ese caso si Pi-pio es un pingüino?

Podemos representar nuestro conocimiento sobre los pingüinos mediante otro condicional standard:

$\forall x$ Pingüino(x) $\rightarrow$ Vuela(x) (2).

El problema es que la representación conjunta de (1) y (2) determina que no pueden existir individuos que sean a la vez pájaro y pingüino.

En la LC, si T es un conjunto de sentencias y P es una sentencia,

si T $\vdash$ P entonces T $\cup$ N $\vdash$ P,

para cualquier conjunto N de sentencias. Si consideramos que:

Pájaro(Pi-pio) (3),
y T es (1), (2) y (3) resulta que T $\vdash$ Vuela(Pi-pio).

Por la monotonía de la consecuencia clásica, si conocemos el nuevo hecho:

Pingüino(Pi-Pio) (4),

tendremos que T $\cup$ {Pingüino(Pi-pio)} $\vdash$ Vuela(Pi-pio), y, a la vez, de (2) y (4) se deduce $\neg$ Vuela(Pi-pío).

Sin embargo, una persona normal, aun aceptando las «reglas» antes señaladas, resolvería la situación aplicando un «principio de predomino de la información más específica», que jerarquiza (2) sobre (1), y descartaría esta última sentencia al opinar sobre Pi-pio, sabiendo que es pingüino.

En esta deducción, hemos aplicado un esquema parecido al ejemplo de Sócrates. ¿Qué es lo que falla? Parecería que no hay dudas sobre que

«Todos los hombres son mortales», mientras que al representar nuestra creencia en la forma «Todos los pájaros vuelan» estamos «forzando» una cuantificación universal, que no es el adecuado reflejo de la expresión del lenguaje corriente «Los pájaros vuelan» (o al menos no lo es en nuestro mundo real, aunque tal vez lo sea en un mundo hipotético alternativo, sin pájaros excepcionales).

Una solución sería representar así nuestro criterio para decidir sobre la capacidad de vuelo de los pájaros:

$\forall$x Pájaro(x) & $\neg$ Pingüino(x) $\rightarrow$ Vuela(x) (5).

El problema es si Pi-pio es avestruz, o tiene el ala rota o las patas amarradas a una piedra, o...

La regla tiene excepciones y no sabemos enumerarlas todas. Sin embargo la seguimos usando a propósito de cualquier pájaro que nos es mencionado, mientras no existan evidencias en contrario (como que es pingüino, que tiene alas rotas, que tiene las patas atadas, etc.).

A estos casos se refería Minsky al indicar la incapacidad de la LC para tratar las propiedades «típicas» de los individuos de un dominio.

Otro ejemplo, inspirado en J. McCarthy, es el siguiente. Supongamos que entre nuestras creencias se encuentra que: «Si el tanque de combustible no está vacío y la batería está cargada, el auto va a arrancar». Pero este condicional no va a ser verdad en un mundo (que podría ser el real), en el que el carburador está roto, o en el cual estén cortados los cables de la batería, o en el cual el tanque, si bien no está vacío, está lleno de agua, o «qualifications», que sería necesario verificar para que se pueda deducir que el auto arranca. Sin embargo, planeamos el día de trabajo sobre la base de verificar sólo las premisas de la regla simple anterior. Si luego se adiciona la información de que el tanque estaba lleno de agua, la conclusión nueva será que el auto no arranca. La conclusión previa es «derrotada» siendo que sus premisas siguen siendo verdaderas.

El llamado *Nixon Diamond,* nos enfrenta a una situación más complicada. Si entre nuestras creencias tenemos que «Normalmente los cuáqueros son pacifistas» y que «Normalmente los republicanos son belicistas», ¿qué podremos afirmar de Nixon, del que sabemos que es cuáquero y republicano? En una formalización «clásica» de este problema, esas dos características no pueden ser verdaderas a la vez en ningún individuo. Si en el caso de los pájaros existía una heurística llamada «principio de predominancias de lo específico», y en el caso del auto otra heurística sobre la relevancia de los distintos factores que condicionan el arranque normal, en este caso, no existen principios intuitivos generales que guíen la respuesta de un agente inteligente. Cada uno podrá preferir una u otra respuesta, según la fortaleza que le asigne a cada creencia, o permanecer agnóstico.

Problemas como los que hemos presentado, han aparecido, desde antes del nacimiento de la IA, en otros campos, notoriamente en el Derecho, la Ética y la Metodología de la Ciencia.

Consideremos el cuerpo legal total en un momento y país dado. Las premisas para que el juez resuelva el pago de un siniestro por parte de una compañía de seguros están todas estipuladas de un modo u otro. Sin embargo el juez resuelve el caso verificando algunas de esas premisas y dejando a la parte demandada el trabajo de presentar pruebas para las otras (por ejemplo una condición de excepcionalidad que invalida el pago). Mientras la compañía no presente pruebas, el juez supone «por defecto» que dichas condiciones de excepcionalidad no existieron, y, si se verifican las premisas principales, concluye que debe pagarse.

En este caso, si bien son conocidas todas las posibles excepciones, su verificación completa resulta muy «costosa». Las conclusiones que surgen son derrotables, en el sentido de que la presentación de pruebas que contraríen la «asignación por defecto», significará su retractación.

En Ética, la discusión sobre el carácter *prima facie* de los principios morales ya trajo antaño conflictos respecto a su formalización. Consideremos el principio moral: «Las promesas deben ser respetadas». Una formalización universal como «Toda promesa debe ser respetada», pierde de vista el carácter *prima facie* del principio, donde las posibles excepciones serán reflejos de conflictos morales (una ruptura de una promesa que sirva para aliviar un sufrimiento). El filósofo W. D. Ross (1927), refiriéndose a este tipo de casos afirma:

> ...Any act that we do contains various elements in virtue of which it falls under various categories. In virtue of being the breaking of a promise, for instance, tends to be wrong; in virtue of being an instance of relieving distress it tends to be right...

Sin embargo, para Ross, la ruptura de la promesa, aun justificada, no elimina el reconocimiento de la obligación «prima facie» de mantener las promesas.

¿Cómo decidir sobre la conveniencia o no de aplicar en cada caso los «principios *prima facie*» o sus excepciones, de forma de sortear el conflicto?

En el campo del Método Científico, en Black, 1935, se afirma que, en general, es imposible explorar cada una de las posibilidades que podrían ser relevantes para la solución de un problema. El único procedimiento factible es tomar la verdad de cierto número de supuestos como dada, y concentrar la atención en el testeo de las hipótesis principales. La decisión acerca de cuáles proposiciones serán consideradas hipótesis principales y cuáles subsidiarias, no puede especificarse en una regla y debe remitirse a un juicio sensato.

La alternativa a una larga y completa, pero muy costosa, si no imposible, descripción de las precondiciones de una regla es basar las conclusiones sólo en información parcial, y rezar para que los factores que han sido ignorados no aparezcan. En ese caso habrá que estar preparados

para que cada tanto se produzcan errores en las conclusiones (el costo de relajar el «control de calidad de las inferencias»). Este problema de equilibrar el monto de conocimiento requerido para hacer una inferencia por un lado, con la exactitud de la inferencia por el otro, es el llamado en IA *qualification problem* (el problema desaparece, por supuesto, si se suponen mundos ideales, de modo de operar como si las excepciones no existieran, pero el sistema así construido no será muy util).

Tomar los atajos e ignorar mucha de la información que es potencialmente relevante, pagando el precio de tener que retraer algunas conclusiones frente a evidencia contradictoria, es el camino que recorren los procedimientos para el razonamiento no monótono.

3. *Lógica y complejidad*

Supongamos que se desea interrogar al sistema, cuyas informaciones almacenadas son $A_1,\ldots,A_n$, sobre si «cree» en B. Si entendemos esto como la pregunta acerca de si B se da en los estados del mundo concebidos por el agente en que se dan $A_1,\ldots,A_n$, la lógica standard nos proporciona un método de prueba puramente sintáctico. La aplicación de este método equivale a hallar alguna secuencia de fórmulas del lenguaje, que termine en B y que esté compuesta sólo por $A_1,\ldots,A_n$ y los axiomas de la lógica subyacente o por fórmulas derivadas de las anteriores de la secuencia por medio de las reglas de inferencia $R_1,\ldots,R_m$.

El primer inconveniente es que, si el lenguaje es el del cálculo de predicados de primer orden, esto no es posible en general, ya que dicho cálculo no es decidible. Existen subconjuntos decidibles de dicho lenguaje, pero incluso en esos casos, considerando que el lenguaje posea negación y disyunción, el problema de la decisión (decidir si una fórmula es o no teorema) resulta computacionalmente intratable. Por lo tanto existirán situaciones en que la respuesta no aparecerá en tiempos razonables, y no se puede prever cuándo estas situaciones van a ocurrir. Esto es preocupante si el robot tiene que actuar en tiempo real (por ejemplo un robot industrial).

La aceleración de los tiempos de proceso por mejoras tecnológicas no resuelve el problema, precisamente porque el peor caso no tiene una cota de tiempo fija. Imaginemos disyunciones del siguiente tipo:

Hecho 1: Adolfo es profesor de IA o de Programación.
Hecho 2: Jorge es profesor de Lingüística o de Sistemas Operativos.

Es claro que tenemos cuatro posibilidades a considerar. Si los hechos aumentan, los casos crecen exponencialmente. Si agregamos procesadores para su cómputo, los tiempos decrecen linealmente. Esto conduce a la siguiente tabla, extraída de Levesque (1986b), donde TIEMPO 1 representa una velocidad de análisis de un millón de posibilidades por segundo, el TIEMPO 2, un millón de millones de posibilidades por segundo, y el TIEMPO 3, un millón de máquinas en paralelo, cada una de ellas analizando un millón de millones de casos por segundo:

Hechos	*Casos*	*Tiempo 1*	*Tiempo 2*	*Tiempo 3*
1	2	—	—	—
2	4	—	—	—
5	32	—	—	—
10	1.024	—	—	—
25	30 millones	30 seg.	—	—
50	10 ** 15	30 años	20 min.	0,001 seg.
100	10 ** 30	10 ** 16 años	30.000 mill. años	30.000 años

Si estamos representando hechos simples por medio de expresiones atómicas o sus negaciones, las disyunciones del tipo ejemplificado arriba van a ser usadas cuando el conocimiento del dominio es incompleto.

Del mismo modo, la negación de un hecho puede sugerir una disyunción de muchos otros, si el dominio es rico en individuos:

$\neg$ Profesor(Raúl), que nos dice que Raúl no es profesor, pero no nos dice qué es Raúl ni quiénes de los otros individuos es profesor,

$\exists$x.Profesor(x) $\wedge$ Dicta(x,IA) que dice que existe al menos un profesor de Inteligencia Artificial, sugiriendo una disyunción sobre los individuos que podrían serlo.

La pobreza de conocimientos acerca del dominio requiere, para ser representada, una mayor riqueza del lenguaje, independientemente de si se utiliza o no un formalismo basado en LC. En este sentido, la Lógica de primer orden permite un alto grado de expresión de estas «indeterminaciones» del conocimiento. El precio a pagar consiste, en cualquier caso, en mayor complejidad computacional.

Empobrecer el lenguaje, al menos si la expresión del problema lo admite, es una forma de restringir la complejidad. Un ejemplo son los lenguajes basados en cláusulas de Horn, en que se evitan la disyunción, la negación y la cuantificación existencial. Esto evita reescrituras complejas de hechos simples, como las derivadas de equivalencias lógicas $(P \equiv (p\&q) \vee (p\& \neg q))$.

Otra alternativa consiste en el «forzar el completamiento» de la información de la BC, mediante ciertos mecanismos *ad-hoc* que se supone que los agentes inteligentes utilizan para actuar frente a incertidumbre (valores por defecto, reglas heurísticas, supuestos de mundo cerrado, etc.). Esto permite llegar a conclusiones que no están implicadas deductivamente por aquello que la BC conoce y que luego pueden ser derrotadas al adquirirse nueva información.

Dentro de estos mecanismos pueden encuadrarse, ahora desde otra motivación, las caracterizaciones prototípicas de los objetos del dominio: «Si es pájaro, entonces vuela» (condiciones necesarias prototípicas del ser pájaro) o «Si vuela y canta, entonces es pájaro» (condiciones suficientes prototípicas).

Consideremos el caso de un robot que, de acuerdo con sus objetivos, decide trasladarse en una habitación desde un punto A hasta otro punto B.

Las modificaciones que esta acción produzca en su BC, le permitirán saber que, luego de dicho traslado, su posición en el escenario en que está operando será B. Pero también necesitará saber qué ocurre con el escenario en su totalidad luego de ese movimiento. En verdad, nada asegura que otros factores presentes en la escena no hayan cambiado mientras el robot ejecutaba su movimiento. Volver a un proceso deliberativo que analice el curso de acción posterior requeriría, entonces, una nueva caracterización de todo el escenario. Esto es muy costoso. Por otra parte no es el modo de actuar inteligente que queremos construir. Luego del movimiento del robot no parece razonable que la pintura del techo haya cambiado de color (¡aunque no es imposible!). Un supuesto «razonable» es que sólo cambia lo mínimo imprescindible de acuerdo con las acciones explicitadas y las proposiciones que describen el escenario. Este supuesto, conocido como *frame axiom,* completa la descripción del escenario en forma no segura, pero permite operar al sistema con mayor eficiencia (evitando considerar todas las alternativas de cambios), a riesgo de cometer errores.

Estos mecanismos de completamiento nos llevan nuevamente al campo del razonamiento no monótono, ya que el agregado de información puede llevar a deshacer inferencias previas. Si bien esto también tiene su costo, si la elección de las suposiciones es razonable, en el sentido de que sea poco probable que luego deban ser levantadas, el balance final de limitar los casos sobre los que se razona será ventajoso. Este también es el caso en que se posee información completa, pero en que el análisis exhaustivo es muy complejo.

Sin embargo, pese a lo atractivo del planteo, los mecanismos de «completamiento» no siempre redundan en una reducción de la complejidad, como se comenta en la próxima sección.

4. *La consistencia*

Otro cuestionamiento a la utilización de la lógica estándar, según las ideas de Minsky, es la condición de consistencia como requisito para desarrollar teorías no triviales (en las que no toda fórmula del lenguaje es derivable de los axiomas de la teoría).

Puede ocurrir que una BC tenga alguna inconsistencia «local», que puede considerarse «irrelevante» en vista del conjunto total de información contenida. La aspiración es poder continuar sacando conclusiones interesantes. Sin embargo, en LC, esta Base sería trivial, en el sentido de no poseer ningún modelo.

En el área de Sistemas Expertos, distintos expertos suelen no coincidir sobre un mismo aspecto del conocimiento del dominio. Si se considera un problema de diagnóstico médico, es natural que, a partir de los mismos síntomas observables, distintos especialistas tengan opiniones conflictivas. En los textos estándar del área se recomienda «evitar» esta situación por medio de algún tipo de arbitraje. Esto no siempre es posible, ni tampoco es posible, en general, «depurar» grandes Bases de Conoci-

miento. Probablemente, en casos como el de la medicina, tampoco sea deseable: conocer la existencia de diagnósticos en conflicto es muy importante para el paciente.

En los últimos tiempos se han desarrollado aplicaciones a IA de las lógicas paraconsistentes. Una lógica es paraconsistente si puede ser la lógica subyacente de teorías inconsistentes, pero no triviales. Estas lógicas fueron propuestas en forma independiente por el lógico polaco Jaskowsky y por el brasileño Newton Da Costa, y desarrolladas ampliamente por éste último, con aplicaciones interesantes a la IA. Entre muchos artículos, se puede consultar Da Costa y Marconi (1989); Da Costa y Subrahmanian (1989) y D'Ottaviano (1990).

El problema de mantener consistencia es central en las formalizaciones del RNM, y produce una complejidad computacional extra. Los conflictos que se desean evitar al usar las inferencias por defecto, como en el caso del *Nixon diamond,* pueden, en un contexto paraconsistente, verse como situaciones normales, causadas por la falta de información completa: «... thinking is a process of resolution of conflicts, by the analysis of evidences. Better saying, thinking is a process of resolution of conflicts, if possible. It may perfectly happen two opposite conclusions having equal rights to be achieved, under the light of available knowledge. Thus contradiction could not be removed, unless by the drastic measure of dismissing both conflicting conclusions. An alternative, not seriously considered so far, (is)... holding both conclusions and keep on reasoning them out, no matter the contradiction, until incoming knowledge may evantually enable a decision...» T. Pequeno (1990).

En Pequeno (1990) se sugieren formalizaciones próximas a la Lógica *Default,* pero con una lógica de base paraconsistente.

5. *Discusión*

Hemos mostrado cómo, a partir de algunos cuestionamientos básicos al uso de la LC en los sistemas de IA, se motivan búsquedas en Lógica que intentan salvar esos problemas. Pero la lista es muy incompleta.

El aprendizaje, aspecto clave en la definición de inteligencia, se intenta formalizar como un proceso inductivo. Esto reaviva el viejo proyecto de fundar una lógica inductiva, que viene desde los albores de la ciencia experimental. El proyecto consistía en ampliar los conocimientos elaborados a partir de los datos experimentales incluyendo algunos que no están totalmente justificados por la verdad de las premisas. Nuevamente, este enfoque determina una noción no monótona de inferencia, ya que nuevos datos (nuevas experiencias de aprendizaje) pueden no encajar en las reglas generalizadoras e invalidar conclusiones previas. Se intentan desarrollar también «lógicas de la analogía», siempre con vistas a inferir consecuencias plausibles a partir de casos previos similares.

La pregunta desde la lógica es: ¿cuál es el control de calidad que le pedimos a estas transformaciones? Obviamente, la respuesta no podría

ser «ninguno», ya que entonces las conclusiones serían totalmente arbitrarias y no podríamos hablar de procesos inferenciales más o menos generales.

Aun relajando el control de calidad de las inferencias, éstas deben tener cierto grado de confianza, en el sentido de que no nos van a defraudar en la mayoría de los casos. Esto sugiere una definición *ad-hoc* de cuándo algo va a ocurrir más frecuentemente que lo opuesto. Si para inferir B a partir de A ya no exijo que B sea verdadero en todas las situaciones en que A es verdadero, quiere decir que existirán situaciones en que $\neg$ B y A son verdaderos y otras en que B y A son verdaderos. Qué será lo más conveniente aceptar en esos casos, dependerá de algún criterio de preferencia entre los A&B estados y los A&$\neg$B estados.

V. INFERENCIAS NO MONÓTONAS

> ...Monotonic logics lack the phenomenon of new information leading to a revision of old conclusions... (McDermott y Doyle).

Los ejemplos de la sección IV sugieren que, dada una colección de items de información, representados como un conjunto P de proposiciones en algún lenguaje lógico, las conclusiones deductivas no son suficientes para satisfacer los requisitos del razonamiento de «sentido común». Esta constatación motivó, a partir de fines de los años 70, el desarrollo de numerosos procedimientos para formalizar el razonamiento no monótono.

En estos procedimientos se establecen reglas de inferencia que permiten «saltar» a conclusiones no establecidas deductivamente a partir de las premisas. Estas conclusiones no son seguras, en el sentido de que las premisas no son condiciones suficientes para su obtención. El agregado de nuevas evidencias puede llevar a la cancelación de inferencias previas.

El tipo de reglas inferenciales utilizadas ya no puede ser las del tipo: Si $P_1,\ldots,P_n$ entonces Q (lógica estándar) porque en ese caso el agregado de nuevas premisas no invalida la conclusión.

Las reglas de inferencia no monótonas son de tipo global y siguen *patterns* como el siguiente:

Si $P_1,\ldots,P_n$, y no se da que $R_1,\ldots R_m$, entonces Q,

donde el añadido de Rj puede invalidar la conclusión.

Lo que se espera de un sistema inferencial no monótono es que, dados P y P' conjuntos (finitos) de sentencias y A una sentencia, si se da que P |~~ A, no se siga que $P \cup P'$ |~~ A, donde |~~ es el símbolo usado para denotar una tal relación de inferencia.

Esta caracterización negativa es muy vaga y abarca demasiados formalismos, motivados a veces desde perspectivas diferentes. Una de ellas

es la del razonamiento «presuntivo» o por defecto, en contraposición a los sistemas basados en estimaciones de probabilidad o plausibilidad. No siempre el razonamiento presuntivo tiene una interpretación estadística. Esta última puede darse en el caso de supuestos como «si me hablan de un pájaro, supondré que vuela», pero no en el de los criterios presuntivos utilizados por los jueces, como «todo acusado es inocente mientras no se demuestre su culpabilidad», ni en el caso de razonamientos «autoepistémicos» como «no tengo hermano mayor, porque si tuviera uno lo sabría».

Los sistemas más conocidos para formalizar el razonamiento por defecto son: la negación por falla de Prolog, caracterizada en Clark (1978), la Lógica *Default,* en Reiter (1980), *Circunscription,* en McCarthy (1980 y 1986), Lógica Modal No Monótona, en McDermott y Doyle (1980), Lógica Autoepistémica, en Moore (1985), y Redes con Herencia, en Touretzky (1986). Cada uno de estos sistemas se presentó con sus propios principios, sin que existiese un marco general en el cual todos pudiesen ser comparados.

La generación de formalismos para reconstruir el RNM, y la búsqueda de sus principios generales ha sido uno de los estímulos más destacados desde la IA sobre el campo de la Lógica, y muchos investigadores de esta última disciplina se han involucrado activamente.

1. *Características de los procedimientos no monótonos*

El antecedente más difundido de los procedimientos no monótonos es la «hipótesis de mundo cerrado» (HMC), que se agrega a la información contenida en una Base de Datos (BD). La HMC estipula que, si una porción atómica de información no se puede extraer de la BD, se supone que vale su negación. El efecto de este supuesto en una BD en que se representan conexiones aéreas, es que, si no se puede obtener de la BD una conexión entre Jujuy y Río Gallegos, es porque no existe tal conexión. En ese caso, se considera que la no existencia de la conexión es una consecuencia de la BD extendida con la HMC. Si luego se agrega a la BD información que permite establecer dicha conexión, la conclusión negativa previa desaparece.

En el caso más general de una Base de Conocimientos (BC) constituida por una colección de cláusulas Horn (por ejemplo, un programa Prolog con negación por falla), este principio se formula así, para cualquier literal positivo «ground» P(t):

(RHMC) Si $BC \nvdash P(t)$, entonces $BC \mid\sim\sim \neg P(t)$,

donde $\mid\sim\sim$ representa la inferencia no monótona inducida por la HMC.

Si analizamos la estructura de la regla (RHMC), hallamos varias características distintivas:

1) la regla es de tipo global, es decir, que la inferencia depende de todas las consecuencias de la BC,

2) la inferencia depende de los conocimientos presentes en la BC, pero también de los ausentes,

3) desde el punto de vista del control de calidad de las inferencias, si bien la regla no es segura (la presencia del conocimiento almacenado en la BC no es suficiente para garantizar la conclusión), sin embargo establece un criterio de resguardo de consistencia en la primera parte de su formulación. Este criterio estipula que $\neg$ P(t) se satisfaga en al menos alguno de los mundos descritos por la BC,

4) satisfechas las garantías, se efectiviza el «salto a las conclusiones» que era el objetivo original. De este modo se incorpora $\neg$ P(t) a las consecuencias ampliadas de la BC (o sea, las consecuencias no monótonas de BC). Los modelos de la BC ampliada serán el subconjunto de los modelos de la BC original en que se satisface $\neg$ P(t). Mas en general, las consecuencias no monótonas de la BC original serán las consecuencias standard de BC $\cup$ HMC(BC), donde HMC(BC) es el conjunto de informaciones agregadas por la (RHMC).

Podemos ejemplificar este mecanismo con nuestro caso canónico:

1) $\forall$x (Pájaro(x)& $\neg$ Anormal(pájaro(x))→Vuela(x))

2) $\forall$x (Pingüino(x)→Anormal(pájaro(x)))

3) Pájaro(Pi-pio),

que nos expresa que «normalmente los pájaros vuelan», «los pingüinos son pájaros anormales» y que «Pi-pio es un pájaro». En principio nada podemos afirmar sobre la capacidad de volar de Pi-pio, dado que no sabemos que sea o no anormal como pájaro. Pero si aplicamos la (RHMC) a Anormal(pájaro(Pi-pio)), podemos concluir no monotónicamente:

4) $\neg$ Anormal(pájaro(Pi-pio)), y por lo tanto,

5) Vuela(Pi-pio).

Si luego se agrega:

6) Pingüino(Pi-pio),

ya no podremos derivar 4) ni 5) de la BC.

Los «criterios de control de calidad» de las inferencias no monótonas suelen tener por objetivo el impedir que la BC se convierta en inconsistente ante la llegada de nueva evidencia (como que Pi-pio es pingüino).

La (RHMC) no puede aplicarse a una BC que no sea formada por cláusulas Horn, porque en ese caso fracasa el «control de calidad» impuesto por la primera parte de la regla. En efecto, si tenemos una BC compuesta sólo por la disyunción P$\vee$Q, la aplicación reiterada de la (RHMC) permite inferir $\neg$ P y $\neg$ Q, y luego $\neg$ (P$\vee$Q), lo que constituye una contradicción con la BC original.

El procedimiento de Circunscripción de McCarthy generaliza la noción de inferencia no monótona basada en la HMC para superar sus limitaciones. Se basa en la idea de seleccionar como modelos «preferidos» de una BC los que posean extensiones minimales de ciertos predicados como por ejemplo Anormal. Esta restricción de los modelos equivale a reforzar la BC, lo que se hace por medio del «Axioma de Circunscripción» (AC), que depende de la BC y de los predicados que

se están circunscribiendo. Las conclusiones por defecto serán las que se obtengan deductivamente de la BC ∪ AC(BC, Anormal), y serán satisfechas en todos los modelos de esta Base ampliada, es decir, en los modelos preferidos seleccionados de la BC original.

En la Lógica *Default* (LD) de Reiter, otro de los formalismos conocidos, se aumentan las conclusiones deductivas de un conjunto de axiomas W, mediante el agregado de reglas de la forma:

A:B/C

Una lectura informal de dicha regla es:

«si A se da en la extensión y no se da ¬ B (es consistente suponer B), entonces infiera C».

Su uso genera extensiones de las consecuencias standard de W.

Una Teoría *Default* es un par ⟨W,D⟩, donde W es un conjunto de fórmulas y D un conjunto de reglas Default. En nuestro ejemplo:

D = {Pájaro(x): Vuela(x)/Vuela(x)}

W = {Pájaro(Pi-pio)}.

Dado que en el contexto de esa teoría no es inconsistente suponer que Pi-pio vuela, la extensión (única) de la teoría incluirá la conclusión (derrotable) Vuela(Pi-pio). No es difícil observar que el mecanismo de la LD, aunque distinto al de Mundo Cerrado y Circunscripción, mantiene las características generales comentadas más arriba. En particular, en cada regla del tipo «normal»:

A:B/B,

la condición «es consistente suponer B» constituye el control de calidad que, si es satisfecho, nos permite pasar de la premisa A a la conclusión por defecto B.

Incluyendo conocimientos sobre pingüinos, obtenemos:

$$D1 = \left\{ \frac{\text{Pájaro(x): Vuela(x)}}{\text{Vuela(x)}} \; ; \; \frac{\text{Pingüino(x):} \neg \text{Vuela(x)}}{\neg \text{Vuela(x)}} \right\}$$

W1 = {Pingüino(Pi-pio), ∀x Pingüino(x) →Pájaro(x)}.

En este caso nos encontramos con dos posibles «extensiones» o escenarios, en uno de los cuales Pi-pio vuela y en el otro no vuela. Si adoptamos una visión escéptica (considerar sólo las consecuencias que surgen en todas las extensiones) nada podemos afirmar sobre esa propiedad de Pi-pio. La razón del comportamiento antiintuitivo es que falta la información de que los pingüinos son pájaros excepcionales respecto al volar (que en el ejemplo de aplicación de la HMC se reflejaba en (2)). Esto nos lleva a modificar las reglas, incluyendo en las mismas las excepciones conocidas:

D2 = {Pájaro(x) & ¬ Pingüino(x): Vuela(x)/Vuela(x);
Pingüino(x): ¬ Vuela(x)/ ¬ Vuela(x)}.

Por un lado esto revela una cierta «fragilidad» en la representación, ante la aparición de nuevas excepciones. Por otra parte, nótese que la regla modificada sigue siendo una regla por defecto, ya que, en el mundo real, existen otros casos de pájaros que no vuelan.

Desde el punto de vista semántico, en la Lógica *Default,* se restringen los modelos de W de acuerdo a las reglas *Default.* En nuestro primer ejemplo, se descartan aquéllos en que Pi-pio no vuela. Los restantes modelos seleccionados de la BC son exactamente los modelos de la extensión. Entre estos últimos puede no estar el mundo real. La inferencia Vuela (Pi-pio), a partir de Pájaro(Pi-pio), puede no ser válida en dicho mundo. El «control» de la regla sólo nos asegura que existe algún mundo compatible con la BC, en que Pi-pio vuela. En el caso de la teoría ⟨W1, D1⟩, existen dos subconjuntos de modelos de la BC en competencia. La modificación de la regla en D2 representa la preferencia por uno de los dos.

En el caso de la HMC, si considero las sentencias (1) a (3), entre todos sus modelos existirán algunos en que Pi-pio es anormal (y no vuela) y otros en que no es anormal (y vuela). La condición de la regla (RHMC) nos asegura que estos últimos existen, y los modelos de las conclusiones extendidas serán aquellos en que la extensión de Anormal es mínima (en este caso, vacía). Si se agrega (4), el panorama cambia, la extensión mínima de Anormal ahora incluye el individuo Pi-pio, y el control de calidad de la (RHMC) impide la aplicación de la regla, ya que, ahora, en todos los mundos compatibles con la nueva BC, Pi-pio es anormal.

La aplicación de las reglas de la Lógica *Default* o de esquemas como Circunscripción sobre alguna propiedad reflejada en la BC, equivale a restringir el conjunto de modelos de la BC según ciertas heurísticas que el constructor de la BC tiene en mente: por ejemplo, que un pájaro, del que no se conoce más información, vuela. Estos modelos preferidos por el diseñador de la BC, serán los modelos de la extensión generada por el procedimiento no monótono. En el «Nixon diamond» la preferencia puede establecerse a partir de una mayor confianza en un supuesto que en otro. Si no es posible establecer esta preferencia, no hay restricción de modelos y el sistema se mantiene «agnóstico» (o presenta las dos alternativas contradictorias).

2. *La lógica de las inferencias no monótonas*

D. Gabbay, en Gabbay (1985), fue el primero en preguntarse, dada una relación |~~ entre enunciados, cuáles serían las propiedades formales que la podrían caracterizar como la relación de inferencia de un sistema no monótono. El punto de partida para este análisis fue la consideración del caso de una relación estándar deductiva ⊢. La respuesta en este caso había sido dada por A. Tarski. Si ⊢ satisface las tres condiciones que siguen (y que están expuestas en una versión finitaria, es decir, considerando sólo conjuntos finitos de premisas), es la relación de inferencia de algún sistema de lógica deductiva.

Reflexividad: $A_1,\ldots,A_n,\ B \vdash B$

$$\text{Cut: } \frac{A_1,\ldots,A_n \vdash X;\ A_1,\ldots,A_n,\ X \vdash B}{A_1,\ldots,A_n \vdash B}$$

Monotonía: $\dfrac{A_1,\ldots,A_n \vdash B}{A_1,\ldots,A_n,\ X \vdash B}$

Los distintos sistemas deductivos conocidos se obtienen agregando diversas propiedades a este conjunto mínimo. En particular, si se desea trabajar con un lenguaje más rico, que contenga las conectivas clásicas, deben agregarse las propiedades que caracterizan a dichas conectivas.

Gabbay propuso, análogamente, unas propiedades mínimas, que debería satisfacer una relación de inferencia no monótona. Estas son Reflexividad, Cut y una forma más débil de monotonía, que fuera bautizada en Makinson (1989) como Monotonía Cautelosa (*cautious monotony):*

Monotonía Cautelosa: $\dfrac{A_1,\ldots,A_n \mid\sim\sim X;\ A_1,\ldots,A_n \mid\sim\sim B}{A_1,\ldots,A_n,\ X \mid\sim\sim B}$

En este contexto, el Cut expresa el hecho de que una conclusión plausible es tan segura como los supuestos en los que está basada, y por lo tanto se puede «acumular» en las premisas. En otras palabras, no hay pérdida de confianza en la cadena de derivaciones plausibles. Esto no ocurre en las inferencias de tipo probabilístico, y es un hecho que este tipo de inferencias no satisface Cut.

Cautious Monotony expresa el hecho de que incorporar una nueva premisa, cuya verdad había sido concluida plausiblemente de los conocimientos previos, no debería invalidar las viejas conclusiones.

Siguiendo a Makinson, se define una relación de inferencia como cumulativa si y sólo si satisface Reflexividad, Cut y Monotonía Cautelosa.

Independientemente, en Shoham (1987), se propuso una teoría de modelos general para las inferencias no monótonas. Sabemos que, en lógica estándar, una proposición B se sigue lógicamente de otra proposición A, y lo notamos:

$A \models B$ si y sólo si B se satisface en todos los modelos de A.

Es inmediato que con dicha definición, por ser los modelos de A&X un subconjunto de los de A, la consecuencia lógica $\models$ es monótona. Shoham sostiene que una noción de consecuencia lógica no monótona puede caracterizarse a partir de algún subconjunto de los modelos de A:

$A \mid\sim\sim B$ si y sólo si B se satisface en los modelos «preferidos» de A.

Dada una lógica estándar L, Shoham construye una Lógica Preferencial $L_{<}$, añadiendo al conjunto de interpretaciones de L, una relación de «preferencia» ($<$) entre ellas. La relación de preferencia es un orden parcial y un mundo V es preferible a otro mundo W, si el agente considera a V como «más normal» que W. Así es posible, dado A, concluir «por defecto» B, si todos los mundos «más normales» entre los A-mundos, también satisfacen B. En otras palabras, B se sigue «por defecto» de A en $L_{<}$, si los B-mundos son un superconjunto de los A-mundos «más normales».

Esta noción de «normalidad» es una generalización de la HMC y Circunscripción, tal como lo muestra el ejemplo de V.1. Allí los mundos

«más normales» eran aquellos modelos de la BC que poseían la mínima extensión de Anormal posible. Si en todos esos mundos «más preferidos», según dicho criterio, Pi-pio no es anormal, entonces concluimos, por defecto, que vuela.

Para Shoham, todo sistema no monótono puede tener una semántica en términos de una relación de preferencia adecuada.

En Makinson (1989) se generaliza la definición de las estructuras de modelos preferenciales. Una estructura M es una terna:

$M = \langle S, \models, \langle \rangle$,

donde S es un conjunto arbitrario no vacío («estados»), ⟨ una relación (de «preferencias») en S y ⊨ una relación entre las sentencias del lenguaje y los elementos de S (de «satisfacibilidad»).

Una estructura de modelo preferencial M induce una relación de inferencia |∼∼ m de la siguiente manera:

A |∼∼ m B si y sólo si, para todo $s \in S$, si $s \models_{\langle} A$, (satisface preferencialmente A según la relación ⟨), entonces $s \models B$, donde

$s \models_{\langle} A$ si y sólo si $s \models A$ y no existe $s' \in S$, con $s' _{\langle} s$ tal que $s' \models_{\langle} A$.

La caracterización más precisa de los elementos del modelo permite definir distintas familias de relaciones de inferencia inducidas, y estudiar las propiedades que poseen.

En particular, la clase de las relaciones cumulativas definidas por Gabbay coincide con las inducidas por las estructuras de modelo preferencial *stoppered* (se dice que $M = \langle S, \models, \langle \rangle$ es *stoppered* si, dado cualquier subconjunto T de S y un elemento t en él, o bien t es ⟨*minimal,* o bien existe t' en T, t' ⟨ t y t' *minimal*).

Con estas herramientas es posible analizar los sistemas de RNM propuestos en la literatura. Esto se realiza en Makinson (1991).

El procedimiento inferencial basado en la HMC y aplicable a cláusulas de Horn extendidas, resulta ser cumulativo. Si se considera S, el conjunto de todas las cláusulas Horn expresables usando un conjunto de símbolos de predicado y símbolos de función primitivos, $\{A_1, A_2, \ldots, A_n\}$ un subconjunto finito de S, que denominaremos A, y B una sentencia, la relación de inferencia inducida por este procedimiento es:

A|∼∼B si y sólo si $A \cup HMC(A) \vdash B$,

donde HMC(A) es el conjunto de los átomos negativos *ground* que se infieren de A por la (RHMC), es decir:

$HMC(A) = \{ \neg P(t) \colon A \nvdash P(t) \}$.

Con esta definición es fácil verificar que |∼∼ cumple las tres propiedades de Gabbay.

Makinson demostró también que la noción de inferencia subyacente en Circunscripción de McCarthy es cumulativa.

Por el contrario, la Lógica *Default* de Reiter, no resulta cumulativa, aun en su versión más «conservadora» (considerando las inferencias plausibles de un conjunto A, con reglas *Default* normales D, como aquellas válidas en todas las extensiones *Default* del mismo).

Consideremos una teoría *Default* con dos reglas:

W_1: {∅} y D: {∅:P/P; P∨Q: ¬P/¬P}.

Las consecuencias no monótonas $C(W_1)$ coinciden con Cn({P}). Consideremos ahora otra teoría con las mismas dos reglas y con:

W_2:{P∨Q}.

Esta teoría tiene tiene dos extensiones, ya que las dos reglas son aplicables, pero no compatibles. Una extensión es $E_1 = Cn(\{P\})$ y la otra es $E_2 = Cn(\{\neg P, Q\})$. Definiendo $C(W2) = E_1 \cap E_2$, es claro que P no pertenece a C(W2).

La situación es, entonces, la siguiente:

$W1 \subseteq W2 \subseteq C(W1)$, pero $C(W1) \not\subseteq C(W2)$, y equivale a la falla de (CM), si definimos A|~~B si $B \in C(A)$

3. *Sistemas no monótonos condicionales*

Un camino parcialmente distinto es el seguido en Kraus, Lehman y Magidor (1990), al caracterizar varias familias de relaciones de consecuencia preferenciales en términos sintácticos y semánticos. Con este propósito aumentan el conjunto de propiedades propuesto por Gabbay, de modo de tratar con un lenguaje que posea las conectivas clásicas. Su objetivo es elucidar las relaciones entre pruebas y modelos, con el objeto de permitir el diseño de procedimientos de decisión que sirvan para realizar inferencias plausibles a partir de Bases de Conocimientos.

Cada familia de relaciones se identifica con cierto tipo de modelos de tipo preferencial. A la vez, todas las relaciones de consecuencia definidas por los distintos modelos de una dada familia están cerradas por un determinado conjunto de reglas de inferencia que las caracterizan sintácticamente.

La relación de inferencia Cumulativa se presenta como la más débil de la familia, y coincide con la propuesta por Gabbay.

Las dos familias que han sido mas tratadas desde entonces son la Preferencial y la Racional.

Una estructura de modelo preferencial KLM es una terna $M = \langle S, l \langle \rangle$, donde S es un conjunto («estados»), $l{:}S \rightarrow U$ una función que asigna a cada estado un «mundo posible» y $\langle$ es un orden parcial estricto en S, que satisface la *smoothness condition* (*stoppered,* en el léxico de Makinson).

La función l permite caracterizar la noción de satisfacción de una proposición A en un estado s, como $l(s) \models A$ en el sentido usual de satisfacción en un mundo.

En Kraus *et al.* se prueba un resultado de representación del sistema P de reglas en términos de los modelos preferenciales definidos arriba. El sistema P está constituido del siguiente modo:

(RW) $\dfrac{\vDash A \rightarrow B,\ C|\!\sim\!\sim A}{C|\!\sim\!\sim B}$ (CM) $\dfrac{A|\!\sim\!\sim B,\ A|\!\sim\!\sim C}{A\&B|\!\sim\!\sim C}$

(LLE) $\dfrac{\vDash A \leftrightarrow B,\ A|\!\sim\!\sim C}{B|\!\sim\!\sim C}$ (REF) $A|\!\sim\!\sim A$

(AND) $\dfrac{A|\!\sim\!\sim B,\ A|\!\sim\!\sim C}{A|\!\sim\!\sim B\&C}$ (OR) $\dfrac{A|\!\sim\!\sim C,\ B|\!\sim\!\sim C}{A \vee B|\!\sim\!\sim C}$

Donde «|~~» es una relación de inferencia:

|~~ ⊆ L × L, y consideramos que el par A, B es equivalente a la conjunción A&B, por lo que la representación finitaria se reduce a una sola fórmula en el lado izquierdo de la relación.

En el sistema P la regla Cut es una regla derivada, lo que permite definir a la operación de inferencia caracterizada por el sistema P como cumulativa:

$$\text{(Cut)}\quad \frac{A|\!\sim\!\sim B,\ A\&B|\!\sim\!\sim C}{A|\!\sim\!\sim C}$$

Una relación de inferencia es racional, si es la relación inducida por un modelo *ranked*. Los modelos *ranked* son modelos preferenciales en los cuales la relación de orden tiene la propiedad *ranked:* existe una función r: S→T, donde T es un conjunto totalmente ordenado por la relación [, tal que s ‹ s' en S si y sólo si r(s) [r(s') en T.

Las relaciones racionales resultan ser las cerradas por el sistema R, que incluye todas las reglas de P, más la siguiente de monotonía racional *(rational monotony):*

$$\text{(RM)}\quad \frac{A|\!\sim\!\sim B,\ A|\!\sim\!\sim \neg C}{A\&C|\!\sim\!\sim B}$$

que expresa el grado más alto de monotonía compatible con una noción de inferencia no monótona.

Llegada a este punto, la relación de inferencia |~~ ha sido embutida en un lenguaje, en el que juega el rol de un conectivo especial, lenguaje gobernado por reglas del tipo deductivo. Las expresiones de este lenguaje son llamadas por KLM «aserciones condicionales». Formalmente, existen notorias similaridades con los sistemas de la Lógica Condicional (LCOND), como algunos de los propuestos, entre otros, en Lewis (1973), y con los sistemas de la obligación condicional, que fueron presentados en von Wright (1971), y en Hansson (1971). Desde el punto de vista sintáctico, el condicional especial de las LCOND, que notaremos ›, no posee la propiedad de refuerzo del antecedente, que sí posee el condicional standard o material, es decir que: de (A›B) no se sigue (A&C›B), lo que le da una característica de «no monotonía» en el plano del lenguaje. Además los axiomas y reglas de inferencia típicos de los sistemas de LCOND muestran una fuerte similaridad con las reglas de P y R.

La siguiente caracterización sintáctica del sistema de Lógica Condicional NP de Delgrande (1987) es ilustrativa al respecto:

(ID) $\vdash A \rangle A$

(CC) $\vdash ((A \rangle B) \& (A \rangle C)) \rightarrow (A \rangle (B \& C))$

(RT) $\vdash A \rangle B \rightarrow ((A \& B) \rangle C) \rightarrow (A \rangle C))$

(CV) $\vdash \neg (A \rangle B) \rightarrow ((A \rangle C) \rightarrow (A \& \neg B \rangle C))$

(CC') $\vdash ((A \rangle C) \& (B \rangle C)) \rightarrow ((A \vee B) \rangle C)$

RCM Si $\vdash B \rightarrow C$, de $A \rangle B$ se sigue $A \rangle C$.

Nótese la similaridad entre los 5 axiomas y las reglas (REF), (AND), (CUT), (RM) y (OR), respectivamente, así como entre la regla RCM y la regla (RW), todas del sistema R.

En realidad las analogías son más profundas, ya que una motivación de dichas lógicas fue el poder representar situaciones en las que el agregado de condiciones «derrote» las conclusiones del condicional, sin generar conflicto (inconsistencias potenciales). En otras palabras, que puedan representarse simultáneamente los condicionales: $A \rangle B$, $A \& C \rangle \neg B$, sin que esto acarree la imposibilidad de A&C.

Por otra parte, las similitudes semánticas se hacen evidentes a partir del trabajo de Shoham. En la LCOND se tiene una medida de similaridad entre mundos posibles. La verdad de un condicional $A \rangle B$ en un mundo de referencia W se establece cuando B es verdadero en ciertos A-mundos «seleccionados», respecto de W. En los sistemas de LCOND contrafácticos clásicos, estos mundos «seleccionados» se caracterizan como los «más próximos» al de referencia en los cuales A es verdadero. En el extremo, si el mundo de referencia W es un A-mundo, el seleccionado será el mismo W.

Sin embargo, hay que señalar como restricción en estas comparaciones que el signo $|\!\sim\sim$, definido por reglas como las presentadas antes, no permite anidamiento (ocurrencias iteradas), ya que está reflejando una noción metalingüística de consecuencia. Por lo tanto, una correspondencia con el condicional especial $\rangle$ de las LCOND, sólo debería tener en cuenta aquellas fórmulas $A \rangle B$ donde ni A ni B contienen a su vez el símbolo $\rangle$ (fórmulas no anidadas o «flat»).

Tampoco aparece en las LNM la noción de mundo de referencia, con lo que la relación de preferencia entre mundos es única en cada modelo preferencial.

A partir de Arlo Costa y Carnota (1989a y 1989b), se comenzaron a establecer formalmente los primeros resultados que conectan sistemas de LCOND y los sistemas de tipo preferencial de las LNM. Posteriormente los mapeos entre LNM y LCOND fueron extendidos en Arlo Costa y Shapiro (1991).

4. *El dilema de las lógicas condicionales*

Antes del desarrollo de los sistemas preferenciales, ya existieron intentos de utilizar sistemas de la LCOND para la formalización del RNM. La

idea era reconstruir afirmaciones prototípicas y «reglas por defecto» mediante el condicional especial. Así, «normalmente los pájaros vuelan», se representaría como:

$\forall x$ Pájaro(x) $>$ Vuela(x).

El filósofo D. Nute, atraído por las propiedades del condicional contrafáctico, se propuso implementar un razonador no monótono mediante un demostrador de teoremas de algún sistema de Lógicas de la obligación condicional, como las presentadas en Hansson (1971), que son un antecedente importante de los sistemas condicionales que buscan la expresión del Razonamiento No Monótono. En el campo deóntico, los mundos «absolutamente normales» son los «mundos ideales» donde todas las obligaciones y deberes son respetados.

Las únicas conclusiones que se pueden extraer con los condicionales derrotables son conclusiones en los mundos ideales o «más normales» (para un análisis más detallado, ver Alchourrón, 1986 y 1991, y Carnota, 1991). Tanto en las LNM preferenciales P y R, como en las LCOND correspondientes es válida esta regla (o teorema) de «*modus ponens* debilitado»:

$$(\text{WMP})\ \frac{\text{True} \mid\!\sim\!\sim A,\ A\mid\!\sim\!\sim B}{\text{True}\mid\!\sim\!\sim B}$$

El (WMP) nos dice que si A es verdadero en los mundos «absolutamente más preferidos» (que no suelen incluir al actual), entonces, de la verdad de la aserción condicional A|~~B se puede inferir que B es verdadero en dichos mundos «ideales». Nada nos dice sobre el mundo actual. El (WMP) sirve para razonar sobre las condiciones ideales, pero no sobre las condiciones reales.

Los sistemas de la LCOND sin *detachment*, resultan ser los que formalmente corresponden a los sistemas de la LNM preferencial, del tipo de P o R, por lo que estos últimos resultan compartir las dificultades inferenciales de los primeros.

En Kraus, Lehman y Magidor (1990) se sugiere la posibilidad de usar sistemas como P para obtener respuestas de una BC. La propuesta es: si se tiene A en un stock de hechos y se deriva, mediante las reglas de P, el condicional A|~~B, a partir de la BC, a la pregunta «¿Es esperable B?» se respondería positivamente. Esto parece equivaler a sostener el «detachment» para |~~. Dado lo informal del comentario es difícil indagar lo que los autores tienen en mente, pero si también es derivable C|~~ $\neg$ B y el stock de hechos contiene A&C, se vuelve al conflicto ya comentado, salvo que se use una lógica especial en los hechos, que no permita derivar ni A ni C de la conjunción.

5. *Las LNM y el detachment de las reglas* Default

¿Cómo funcionan los formalismos conocidos para extraer conclusiones por defecto?

Como se vio al inicio de esta sección, los procedimientos del RNM poseen reglas para autorizar o bloquear el *detachment* de los condicionales *Default,* de acuerdo a determinados criterios (por ejemplo criterios de consistencia). Nuevas evidencias incorporadas a la BC pueden provocar el bloqueo del *detachment* de ciertos condicionales *Default* (para evitar una inconsistencia en la BC), por lo que no se siguen derivando consecuencias previamente establecidas.

Repasemos el funcionamiento de nuestro ejemplo:

1) ∀x (Pájaro(x)& ¬ Anormal(aspecto1(x)))→Vuela(x)

2) Pájaro(Pi-pio).

Con la (RHMC) aplicada a Anormal (aspecto1(Pi-pio)), donde aspecto1 define «ser anormal como pájaro respecto al volar», obtenemos Vuela(Pi-pio). De entre todos los modelos de la BC hemos preferido los que minimizan la extensión de Anormal. El mundo real puede no estar entre ellos, pero al menos sabemos que no es un conjunto vacío.

Si ahora sabemos que Pi-pio es pingüino y que, normalmente, los pingüinos no vuelan, agregamos a la BC:

3) ∀x (Pingüino& ¬ Anormal(aspecto2(x)))→ ¬ Vuela(x)

4) Pingüino(Pi-pio).

Si aplicamos (RHMC) nuevamente, nos encontramos en la misma situación anómala señalada para el caso de la LC. Si agregamos:

5) ∀x (Pingüino(x)→Anormal(aspecto1(x))), la aplicación de la (RHMC) ya no permite inferir Vuela(Pi-pio).

Esto restringe los modelos de la BC a aquellos donde vale Anormal(aspecto1(Pi-pio)). En este punto no podemos inferir nada sobre Pi-pio. Aquí interviene entonces la (RHMC), que infiere por defecto:

¬ Anormal(aspecto2(Pi-pio)), garantizando, como «control de calidad», que exista al menos algún modelo de la BC en que dicha consecuencia se satisfaga (y donde Pi-pio no vuela). Evidentemente no tenemos certidumbre alguna de que las conclusiones por defecto sean valederas en el mundo real. Sólo tenemos la garantía de que, si son extraídas, no es imposible que se verifiquen en dicho mundo. La autorización del *detachment* del condicional por defecto consiste en la aceptación —provisoria— del consecuente, dado el antecedente y provista la garantía de consistencia.

La restricción de los modelos equivale a afirmar en el antecedente del condicional por defecto la negación de todas las excepciones aun no conocidas.

Al especificarse las propiedades de un operador |~~ mediante reglas al estilo de los sistemas P o R, se establecen las condiciones de derivación de condicionales a partir de condicionales, pero no se determinan los mecanismos concretos que controlan el «salto a las conclusiones». En ese sentido son postulados generales que caracterizan completamente una clase de relaciones de inferencia, pero no una relación de inferencia concreta. Cada modelo concreto de esos postulados es una LNM, en la que, bajo ciertos resguardos, se afirma la verdad por defecto del consecuente.

6. *Lógica no monótona y revisión de creencias*

La teoría del cambio de creencias trata de la dinámica de los estados de creencias, con el objetivo de modelizar las actualizaciones de los estados de creencias de un agente o de un sistema de computación, como resultado de recibir nueva información. Existen varios tipos de cambios de creencias. El más simple es el que surge por el aprendizaje de algo nuevo y es conocido como expansión. A veces, sin embargo, estas nuevas evidencias contradicen creencias previamente aceptadas, lo que lleva a una revisión del estado de creencias con vistas a mantener su consistencia. Esta revisión requiere la eliminación de viejas creencias. Otras veces el descubrimiento de que las razones para sostener una creencia han desaparecido conduce a una contracción del estado de creencias. Una revisión de un conjunto de creencias K, como resultado del aprendizaje de una evidencia A, puede ser considerada como la sucesión de una contracción de dicho conjunto por la negación de A, y luego el agregado (por expansión) de A. El primer paso asegura que la incorporación de la nueva evidencia no provocará inconsistencia en K. Una operación de contracción (y por ende una revisión) no es sencilla de definir: dado un conjunto K y una sentencia C, existen varias formas de eliminar sentencias que puedan implicar C. Si incorporamos, como criterio de racionalidad, que la operación redunde en la menor pérdida de información posible, una manera informal de visualizar una contracción de K por una creencia C, es en términos de la familia de los subconjuntos maximales de K que no implican C, que se nota $K \perp C$.

Alchourrón, Gardenfors y Makinson desarrollaron una teoría del cambio racional de creencias, presentando construcciones explícitas de las operaciones de cambio (en particular, las basadas en los subconjuntos maximales citados), así como postulados que dichas operaciones deberían cumplir. Los dos enfoques son conectados en Alchourrón, Gardenfors, Makinson (1985) a través de teoremas de representación.

Los postulados AGM de revisión racional de creencias son ocho, seis de los cuales son denominados básicos y dos complementarios. Si denotamos con * la función de revisión, y consideramos conjuntos de creencias cerrados por consecuencia lógica clásica (teorías), siguiendo a Gardenfors (1988), ellos son:

K*1) Si K es una teoría y A una sentencia, K^*A es una teoría.
K*2) $A \in K^*A$.
K*3) $K^*A \subseteq K^+A$ (la expansión de K por A).
K*4) Si $\neg A \notin K$, entonces $K^+A \subseteq K^*A$.
K*5) $K^*A = K$falso si y sólo si $\vdash \neg A$ (donde Kfalso denota el conjunto de creencias inconsistente).
K*6) Si $\vdash A \leftrightarrow B$ entonces $K^*A = K^*B$.
K*7) $K^*(A \& B) \subseteq (K^*A)^+B$.
K*8) Si $\neg B \notin K^*A$, entonces $(K^*A)^+B \subseteq K^*(A \& B)$.

Consideremos un ejemplo en el que un estado de creencias de un sistema está reflejado por una BC y sus consecuencias lógicas. Esta situación es la normal en la práctica, y se dice que BC es una base del conjunto de creencias K. Las funciones que efectivizan cambios en teorías apelando a cambios en sus bases, son llamadas «revisiones de bases» y están siendo estudiadas entre los investigadores de IA.

Sea, en una situación dada, la BC constituida por:

1) ∀x (Pájaro(x)→Vuela(x))

2) ∀x (Pingüino(x)→ ¬ Vuela(x))

3) Pájaro(Pi-pio).

Entre las creencias implícitas del sistema estarán:

4) Vuela(Pi-pio), y

5) ¬ Pingüino(Pi-pio).

Si ahora obtenemos una nueva evidencia:

6) Pingüino(Pi-pio), para acomodar (6) a la BC, manteniendo la consistencia, debemos eliminar alguna creencia previa, entre 1), 2) ó 3). La elección de qué creencias se eliminan (o de qué subconjunto se prefiere conservar) dependerá de algún criterio de preferencia entre las creencias (o entre los subconjuntos maximales que no implican (5)).

Entre las funciones de contracción propuestas en Alchourrón, Gardenfors y Makinson (1985), la «partial meet contraction» de K por una sentencia C, se define como la intersección de una subfamilia de K ⊥ C. Esta subfamilia es elegida por medio de una función de selección S, que, si bien puede ser arbitraria, es razonable de suponer que escoja los subconjuntos «mejores» de un cierto orden:

$S\ (K \perp C) = \{K' \in K \perp C : K'' \leq K' \text{ para todo } K'' \in K \perp C\}$.

La contracción K^-_C resulta luego: $K^-_C = \cap S(K \perp C)$.

De acuerdo al criterio indicado antes, la revisión de K por A, notada K*A,

(*

En Makinson y Gardenfors (1990), se sugiere un método de traducción entre postulados de revisión de creencias y propiedades de las LNM. La idea básica es ver una expresión de la forma:

B ∈ K*A,

K como conjunto de hipótesis (o expectativas por defecto) auxiliares. A la inversa, una expresión de la forma:

A|~~ B de una LNM, se traduce a una de la forma B ∈ K*A

sión, donde K es introducido como un conjunto de creencias fijo. La forma de traducción es, entonces:

A|~~B si y sólo si B ∈ K*A

Usando esta receta, es posible traducir los 8 postulados de revisión de AGM en reglas que definen propiedades del operador |~~ , en particular, válidas casi todas en el sistema R. A la inversa, los distintos postulados para las LNM se traducen en condiciones de cambio de creencias que son derivadas de los 8 postulados de la revisión AGM. Por ejemplo, la monotonía cautelosa (CM), se traduce en:

Si $B \in K^*A$ y $C \in K^*A$, entonces $B \in K^*A\&C$, que se deriva de los postulados de revisión.

Posteriormente, en Gardenfors y Makinson (1991), se proponen diversos formalismos, «basados en expectativas» para generar procedimientos de inferencias no monótonas. Estos formalismos se inspiran directamente en distintos modelos de la teoría de cambio de creencias, y pueden verse como generalizaciones del trabajo de Poole (1988). Una operación de inferencia basada en conjuntos de expectativas puede definirse informalmente así:

A implica no monotónicamente B si y sólo si B se sigue lógicamente de A junto con tantos elementos como sea posible del conjunto fijo K de expectativas, con la condición de que sean compatibles con A.

Más formalmente, dado un conjunto de expectativas K no vacío y una función de selección S del tipo de la mencionada más arriba, la operación de inferencia C se define como:

$$C[K,S]\ (A) = n\{Cn(\{A\} \cup K' : K' \in S(K \perp \neg A)\}.$$

Esta caracterización es la misma de (*), donde se define una revisión de K por A basada en una «partial meet contraction function». Lo interesante es que si S selecciona los mejores subconjuntos según un orden transitivo, C[K,S] satisface todas las reglas del sistema R (así como la respectiva revisión cumple los 8 postulados), más una regla adicional que no todos los sistemas R cumplen, llamada «Preservación de Consistencia» y que es la traducción del K^*5 (si $A|\sim\sim$ falso, entonces $A \vdash$ falso).

Del punto de vista semántico, es fácil ver que hay una relación uno a uno entre los A-mundos y los subconjuntos de la familia $K \perp \neg A$, por lo que la aplicación a esta última de una función de selección que elige según un orden transitivo, tiene las mismas características de una selección de mundos «preferidos» en una lógica preferencial del tipo de la de Shoham.

Llegado a este punto, se evidencia una correspondencia formal muy sólida entre cierto tipo de revisión de creencias (la denominada AGM) y las relaciones de inferencia no monótonas preferenciales más «fuertes» (más próximas a la LC).

En una inferencia no monótona basada en expectativas se considera un conjunto K fijo de hipótesis y se extraen del mismo conclusiones por defecto (que mantienen el control de consistencia al restringirse a las consecuencias clásicas de sólo un subconjunto de K). En revisión, las conclusiones son siempre deductivas a partir de un nuevo conjunto K', resultante de la previa contracción de las premisas. En ambos casos, dado el mismo criterio de «preferencia» los resultados (las conclusiones extraidas) son los mismos.

Esta vinculación complementa y confirma resultados previos obtenidos en Alchourrón y Makinson (1981) que, en su momento, por el aislamiento mutuo de los campos de la Lógica y la IA no tuvo repercusión en este último.

Una interpretación epistemológica de esta correspondencia, sugerida por Gardenfars y Makinson (1991), es pensar el conjunto de creencias

K en dos fases. Mientras se lo está utilizando, sus elementos son *full beliefs* y se extraen las consecuencias deductivas de los mismos. Pero tan pronto se procede a su revisión, estos elementos son cuestionados, por lo que pierden su status de *full beliefs,* para convertirse en expectativas o hipótesis sobre el dominio, algunas de las cuales deberán ser descartadas con el fin de introducir creencias nuevas, preservando la consistencia del conjunto.

7. *Inferencias no monótonas y complejidad*

En la sección 4 el problema de la complejidad computacional apareció vinculada al tema del RNM, a partir del uso de supuestos, como el «frame axiom», para llenar lagunas del conocimiento. La hipótesis subyacente, tal como se presenta en Levesque (1988), es la siguiente:

> ...The deviations from classical logic that will be necessary to ensure the tractability of reasoning stand in very close correspondence to the deviations from logic that we would have to make anyway to be psychologically realistic. If we look at the kinds of mistakes people make, the kinds of problems people run into, and the corners that are cut to get around them, we will find modifications to classical logic that ensure the computational tractability of the associated thinking...

Uno de los caminos sugeridos para el completamiento de las BC es el de las inferencias por defecto. Como ya se discutió, este camino puede verse como una restricción en el análisis de los modelos sólo a aquellos «preferidos», en algún sentido, por el diseñador del sistema (precisamente aquellos donde no se verifican ciertas excepciones).

El efecto de esta restricción es que se deben analizar menos modelos con la consiguiente simplificación de los procesos de decisión.

Pero el reflejo de esta restricción en la teoría de prueba es inverso. Ahora las inferencias son globales y deben realizar «controles» (usualmente de consistencia), para justificar los «saltos a las conclusiones» no deductivos.

La simplificación del proceso de decisión lleva a la pérdida de la «natural computabilidad» de las reglas de inferencia deductivas.

Si bien a la hora de decidir si B se sigue de $A_1,...,A_n$, ya no debemos explorar exhaustivamente todos los factores en juego (todos los modelos de $A_1,...,A_n$), en cambio trasladamos a las reglas «no monótonas» un problema de decisión aún más complejo como es el de probar que una sentencia no se deriva de otras (problema ni siquiera semidecidible en el caso general de primer orden). Es claro que muchas veces la motivación de dicha inferencia no monótona es la falta de conocimiento y la imposibilidad de inferencia alguna (costosa o no). Pero el problema del test de consistencia ha hecho que los formalismos de raciocinio no monótono carezcan, en general, de implementaciones efectivas, salvo para casos particulares.

Distinguiendo las dos etapas del proceso, es posible extraer conclusiones por defecto en forma rápida. El costo computacional está en los controles de calidad.

Este resultado es paradojal si con el «salto a las conclusiones» se intenta imitar el «razonar de la gente», cuando evita el análisis de todas las circunstancias posibles.

Como se afirmó en II.2., el abordaje de problemas característico de la IA consiste en buscar heurísticas que eviten los análisis de todas las posibles alternativas. De hecho estas heurísticas son «inferencias por defecto *ad hoc*». El problema que detectamos es, entonces, el fracaso en construir marcos formales generales de estas inferencias, que sean «naturalmente computables».

En la práctica, sin embargo, existen casos particulares en los que los formalismos son implementables de modo que realicen inferencias en tiempos razonables.

Estos casos particulares surgen de imponer restricciones al lenguaje en que se expresan las premisas del razonamiento o al tipo de teorías expresables. Ejemplos de esto son las restricciones que se hacen para lograr computar ciertos casos de circunscripción en Gelfond y Lifschitz (1988), o las que se hacen en Shoham (1988), sobre la expresividad de las teorías representadas, o el éxito del procedimiento de negación por falla en programas Prolog, o el uso de valores por defecto en redes semánticas representando taxonomías.

En su trabajo, Shoham demuestra que ciertas restricciones expresivas en las teorías formuladas llevan a restringir el análisis a un único modelo. Una de las restricciones consiste en no permitir reglas *Default* con efectos opuestos y donde las premisas puedan ser consistentes (tener instancias comunes). En el caso conocido del *Nixon diamond:*

los cuáqueros son pacifistas,
los republicanos son no pacifistas,

resulta claro que las premisas son consistentes y ese tipo de situaciones no pueden ser expresadas para asegurar condiciones de computabilidad. En particular, esta restricción evita el tener que establecer criterios de preferencias y tener que optar entre distintos conjuntos de *Defaults* consistentes.

Lo antedicho sugiere la idea de que el factor de complejidad sólo puede ser resuelto por una combinación de inferencias *Default* y restricciones expresivas. El proceso de control para poder aplicar una regla no monótona, es naturalmente más tratable en lenguajes pobres (por ejemplo que sólo admiten cláusulas de Horn). En lenguajes más ricos expresivamente, las inferencias no monótonas conocidas constituyen operaciones de alta complejidad.

8. *La pragmática de las inferencias no monótonas*

¿Cómo deben entenderse estos mecanismos o reglas de «salto a las conclusiones»? ¿Son reglas de inferencia clásicas?

Por un lado, una regla de inferencia lógica se caracteriza por interactuar con los conectivos del lenguaje al margen de todo tipo de denota-

ciones. Una regla que afirma $\neg(\neg a) \leftrightarrow a$, no tiene «contenido» extra, ligado a una teoría de un dominio particular. En ese sentido, podemos decir que es «vacua» de contenido.

Las hipótesis de completamiento, de uso corriente en IA, tales como el *Frame Axiom,* la HMC, etc. suponen una gran cantidad de conocimiento específico. Aceptar $\neg P(t)$ si no es derivable P(t) no es una decisión trivial. La justificación de estas hipótesis está en supuestos heurísticos sobre el comportamiento del mundo (o de ciertos dominios o contextos).

Las reglas *Default* tampoco son generales. En realidad, son sustantivas, en el sentido de que proveen conocimiento específico de lo que se espera que ocurra con ciertas propiedades de ciertos individuos en ciertos dominios.

Si proponemos un *Default* que afirme:

«Si una casa es habitable entonces está calefaccionada»,

resultará muy razonable en Escandinavia, pero muy irrazonable en zonas tropicales.

Esto es así, ya que la razón de la adopción de una hipótesis derrotable es económica. Se espera que, a la larga, si el *Default* esta bien elegido, la mayoría de las inferencias será correcta. El caso es que el concepto de «bien elegido» depende de factores externos —y dependientes del dominio— como el daño que puede hacer una inferencia errónea. Si el 5% de los pájaros no vuela, adoptar el supuesto «Los pájaros vuelan» es razonable. Si el 5% de la gente que anda por la calle tiene la costumbre de dar puñaladas en la espalda, el *Default* «si se cruza una persona por la calle, no es preciso cuidarse la espalda» es peligroso (si el porcentaje cae a 0,000005%, ya sería aceptable).

Este mismo criterio pragmático es el que guía la utilización de normas presuntivas en el sistema jurídico. Un principio que afirma: «Si alguien falta de su domicilio y no da noticias por 5 años, se lo considera, a todos los efectos legales, como fallecido» es adoptado por la justicia considerando que el margen de error será muy bajo (esta vez en base a consideraciones referidas a las normas de convivencia social, y no consideraciones estadísticas) y el beneficio de resolver cuestiones legales trabadas es muy alto.

Los formalismos basados en ordenamientos de sentencias o de conjuntos de estados, en cada caso apelan a criterios pragmáticos (considérese lo que implica en este sentido «preferir» una conclusión u otra en el problema de *Nixon diamond*). Si bien se pueden señalar sus propiedades generales, dadas las propiedades del orden subyacente, cada lógica preferencial concreta es la que se construye a partir de un orden específico. En cada modelo preferencial, el orden particular $<$ refleja un conocimiento del dominio (una intuición del diseñador acerca de cuáles son los «modelos preferidos» en una aplicación de IA). En realidad, cada vez se están considerando teorías parciales de los dominios representados, teorías que engendran los casos particulares de reglas, meta-reglas, ordenamientos, etc. en que se sustentan las inferencias no monótonas.

En resumen, podemos decir que el RNM se inscribe dentro del razonamiento pragmático, en el sentido en que es particular a un contexto, en el cual la información es limitada, y que emplea criterios heurísticos para arribar a conclusiones razonables. Sus reglas son, en realidad, metareglas para razonar sobre dichos contextos particulares. La monotonía es una propiedad característicamente libre de contexto, y por eso no puede caber en este tipo de raciocinio.

Cada LNM particular esta definida en una sola estructura de modelo y las inferencias en este modelo, que podríamos llamar inferencias pragmáticas, se caracterizan por la verdad del antecedente en sólo un subconjunto preferido pragmáticamente de los estados del modelo. Estas restricciones hacen que tal vez sea más adecuado hablar de «Procedimientos Inferenciales No Monotónicos basados en Lógica» y que el problema de la nomonotonía, más que un problema con la lógica sea un problema acerca de cómo la lógica es usada.

VI. LA LÓGICA EN LA INTELIGENCIA ARTIFICIAL (A MODO DE CONCLUSIÓN)

> ...It may appear that logical proof is being opposed to reasoning. The correct view seems to be that logical proof is a tool used in reasoning... (D. Israel).

La pregunta que formulamos al final de la sección III era: ¿es el proceso de elaboración de la información del «robot» un proceso deductivo? De lo discutido hasta aquí, parece evidente que no. Los objetivos de la IA no se satisfacen reduciendo el proceso de elaboración de información de un sistema inteligente a un demostrador de teoremas. Esta evidencia ha promovido en la IA y en la Lógica, desarrollos de lógicas no clásicas. Los más característicos han sido los motivados por el Razonamiento No Monótono. Sin embargo, los formalismos lógicos para el RNM no han resultado satisfactorios, hasta ahora, salvo en casos particulares. A partir de estas dificultades, cabe preguntarse: ¿Cuál es el lugar de la lógica en los procesos de reconstrucción formal del «raciocinio inteligente»?

Creemos que la lógica retiene al menos dos roles de importancia, uno al interior y otro al exterior de los procesos deliberativos de los sistemas inteligentes.

El primero surge a partir de entender las diferencias entre razonar y deducir.

Como hemos visto, el razonar incluye el deducir, pero requiere también ir más allá de lo absolutamente seguro. Es un fenómeno global (y no local, como la inferencia deductiva) y debe tener en cuenta juicios acerca de la relevancia y los pesos de evidencia de los argumentos en juego. Puede juzgar que las evidencias no son suficientes y solicitar más información. Puede (y debe) eliminar viejas creencias, a la luz de nuevas evi-

dencias, de acuerdo a criterios racionales determinados para cada caso. Supongamos que aceptamos una sentencia de la forma «si P entonces Q» y aceptamos el antecedente. ¿Nos obligaría este hecho a aceptar Q? No necesariamente. Tal vez tengamos muy buenas razones «de jerarquía superior» para creer no-Q. En todo caso esto no llevará a rever nuestra creencia en el condicional o en el antecedente.

Eliminar viejas creencias no es en manera alguna ilógico, especialmente si ellas implican conflictos con las que ahora tenemos buenas razones para sostener. El punto es que no debemos esperar que sea la lógica la que nos diga qué retener y qué eliminar, ni que nos diga que hacer cuando —gracias a su ayuda— descubrimos que poseemos creencias inconsistentes.

La deducción lógica, lejos de estar enfrentada al razonamiento, debe verse como una herramienta usada en el proceso de razonar. Los criterios de calidad de la lógica son un punto de referencia sólido para evaluar la confiabilidad de las reglas del razonar.

Razonar es más próximo a revisar creencias. En la base de un sistema de revisión racional de creencias, tendremos alguna lógica, aunque luego el sistema inteligente decida las acciones a tomar en base a un esquema de preferencias de tipo pragmático. Las reglas de inferencia deductivas sirven para explicitar el contenido informativo de la Base de Creencias, pero no alcanzan para determinar las políticas racionales de transformación de dicha Base, que den cuenta de los procesos de «aprendizaje», a partir de las interacciones entre el robot y su medio.

El paradigma del razonamiento de sentido común ha sido, en la última década, el RNM. Hemos visto que es posible entender estas inferencias no deductivas en términos de una combinación de inferencias deductivas y revisión de creencias, evitando la proliferación de nuevas lógicas, que muchas veces llevan a resultados antiintuitivos y/o poco operativos.

En segundo lugar, la lógica tiene un rol descriptivo y, en cierto modo normativo, respecto de los mecanismos implementados en los sistemas de IA.

Aun cuando descartemos la Lógica como simbolismo de representación y elaboración de los datos que el sistema posee, ella es una herramienta adecuada para dar cuenta de los procesos que el sistema realiza en términos más confiables que una descripción computacional.

Frente al argumento de que la lógica es «demasiado prolija» para atacar problemas inherentemente complejos y poco claros —los procesos cognitivos—, hay que coincidir en que cualquier modelo que pretenda echar luz sobre los fenómenos del razonamiento, tiene que poder ser entendido claramente. De lo contrario, dado que el objeto modelado es de por sí poco conocido y la conducta del modelo no es totalmente clara, poco es lo que se podrá concluir o resolver. «In extremis» dicha argumentación contra la lógica es una argumentación contra todo rigor.

Puede argumentarse que la comprensión de los alcances del modelo también puede hacerse en alguna teoría matemática. Esto es cierto. Sin embargo, considerando que:

a) asumimos la restricción de la representación de conocimientos declarativa, que implica una estructura proposicional (aunque no necesariamente en un lenguaje de la lógica), donde una pieza de estructura del formalismo representará una aserción (conocimiento o creencia) acerca del dominio, y

b) dada una representación declarativa (red semántica, «frame», estructura *ad hoc* para posiciones del ajedrez o cualquier otra), interesa determinar exactamente qué conocimiento está siendo representado (no sólo explícitamente, lo que finalmente se reduce a una enumeración, sino implícitamente) y caracterizar, mediante alguna «teoría de la verdad», cuán confiables son los conocimientos implícitos que el sistema puede inferir de los explícitos, en términos del universo que queremos modelizar, entonces la lógica parece la herramienta más adecuada para analizar y comprender el comportamiento del sistema.

Este enfoque es el aplicado por Brachman y Levesque (1984), a la caracterización funcional de un esquema de Representación de Conocimiento (RC). En esta visión, no interesa el detalle de cómo está construido el sistema de RC o qué estrategias usa para ser eficiente. Lo que importa es lo que sabe del mundo (en términos de sus creencias básicas y de su capacidad de derivar de ellas otras creencias). De hecho Makinson ha utilizado la lógica con ese sentido metateórico al estudiar las propiedades de los formalismos del RNM. La lógica es también usada en Balkenius y Gardenfors (1990) para caracterizar el poder inferencial de ciertos tipos de redes neuronales.

En síntesis, un rol fundamental de la lógica en IA es como herramienta para el análisis del contenido de conocimiento involucrado en la Base de Conocimientos (KB) del robot, antes que para reconstruir el modo de razonar de seres inteligentes. Es decir, que la lógica es el marco adecuado para analizar el sentido de las expresiones que aparecen en los formalismos de representación y para juzgar la validez de las inferencias, independientemente de que los lenguajes lógicos sean, en sí mismos, adecuados formalismos de representación, y de que la aplicación de reglas de inferencia deductivas a fórmulas lógicas sea un buen método para reconstruir el razonar de sentido común.

BIBLIOGRAFÍA

AI (1991), Número especial dedicado a los Fundamentos de IA: *Artificial Intelligence, 47.*

Alchourrón, C. y Makinson, D. (1981), «Hierarchies of Regulations and their Logic», en R. Hilpinen (ed.), *New Studies in Deontic Logic,* 125-148, Reidel, Dordrecht.

Alchourrón, C., Gardenfors, P. y Makinson, D. (1985), «On the Logic of Theory Change: Partial Meet Contraction and Revision Functions»: *The Journal of Symbolic Logic, 50,* 510-530.

Alchourrón, C. (1986), «Conditionality and the Representation of Legal Norms», en *Automated Analysis of Legal Texts: Logic, Information, Law,* Elsevier North-Holland.

Alchourrón, C. (1991), «Philosophical Foundations of Deontic Logic and its Practical Applications in Computational Contexts», manuscrito, Universidad de Buenos Aires.
Arlo Costa, H. y Carnota, R. (1989a), «Non Monotonic Preferential Logic and Conditional Logic»: *Proceedings IX Cong. Internacional Soc. Chilena de Computación,* Santiago.
Arlo Costa, H. y Carnota, R. (1989b), «Non Monotonic Logics: Consequence Relations and Conditional Operators»: *Anales VI SBIA*, Rio de Janeiro.
Arlo Costa, H. y Shapiro, S. (1991), «Maps between Non Monotonic and Conditional Logics», manuscrito, Department Philosophy, Columbia University, New York.
Balkenius, C. y Gardenfors, P. (1990), «Non Monotonic Inferences in Neural Networks», en Alles, Fikes y Sandewall (eds.), *Proceedings of the Second Int. Conf. on Principles of Knowledge Representation and Reasoning,* Morgan-Kauffman, San Mateo, Calif.
Birnbaum, L. (1991), «Rigor Mortis: a response to Nilsson's "Logic and Artificial Intelligence"»: *Artificial Intelligence, 47,* 31-56.
Black, M. (1935), *Critical Thinking,* Prentice-Hall Inc., New York.
Brachman, R. y Levesque, H. (1984), «A Fundamental Tradeoff in Knowledge Representation and Reasoning», en Levesque y Brachman (eds.), *Readings in Knowledge Representation,* Morgan-Kauffman, Los Altos.
Carnota, R. (1991), «Condicionales Derrotables e Inferencias No Monótonas», manuscrito, Departamento Ciencias Computación, Universidad de Buenos Aires.
Clark, R. (1978), «Negation as Failure», en Gallaire y Minker (eds.), *Logic and Data Bases,* Plenum, New York.
Cohen, M. y Nagel, E. (1957), *An Introduction to Logic and Scientific Method,* Routledge and Kegan Paul, London.
Da Costa, N. y Marconi, D. (1989), «An Overview of Paraconsistent Logic in the 80's»: *The Journal of Non Classical Logic, 1*, vol. 6, 5-32.
Da Costa, N. y Subrahmanian, V. S. (1989), «Paraconsistent Logics as a formalism for reasoning about inconsistent knowledge bases»: *Artificial Intelligence in Medicine, 1,* 167-174.
Dedalus (1987), *A special issue in Artificial Intelligence,* Pub. of the Academy of Sciences of New York.
Delgrande, J. P. (1987), «A First Order Conditional Logic for Prototypical Properties»: *Artificial Intelligence, 33* (1).
D'Ottaviano,
The Journal of Non Classical Logic, 1-2, vol. 7, 89-152.
Doyle, J. (1989), «Rational Control of Reasoning in Artificial Intelligence», en Fuhrmann and Morreau (eds.), *The Logic of Theory Change,* Springer Verlag, Lect. Notes in Art. Int., 465, 19-48, Berlin.
Gabbay, D. (1985), «Theoretical foundations for Non Monotonic Reasoning in Expert Systems», en *Logics and Models of Concurrent Systems,* Springer Verlag.
Gardenfors, P. (1988), *Knowledge in Flux: Modeling the Dynamics of Epistemic System,* The MIT Press-Bradford Books, Cambridge, Mass.
Gardenfors, P. y Makinson, D. (1991), «Non Monotonic Inference Based on Expectations», manuscrito, Dpt. Philosophy, Lundt University.
Gelfond, M. y Lifschitz, V. (1988), «Compiling Circunscriptive Theories into Logic Programs», en *Proceedings AAAI-88,* Morgan-Kauffman, Los Altos, Calif.
Ginsberg, M. (1988), «Introduction to Readings in Non Monotonic Reasoning», en Ginsberg (ed.), Morgan-Kauffman.
Hansson, B. (1971), «An Analysis of some Deontic Logics», en R. Hilpinen (ed.), *Deontic Logic: Introductory and Systematic Readings,* D. Reidel, Dordrecht.
Kraus, S., Lehmann, D. y Magidor, M. (1990), «Nonmonotonic Reasoning, Preferential Models and Cumulative Logics»: *Artificial Intelligence, 44.*

Levesque, H. (1986a), «Knowledge Representation and Reasoning»: *Annual Reviews of Computer Science, 1,* 255-287.
Levesque, H. (1986b), «Making Believers Out of Computers»: *Artificial Intelligence, 30.*
Levesque, H. (1988), «Logic and the Complexity of Reasoning»: *Journal of Philosophical Logic, 17,* 355-389.
Lewis, D. (1973), *Counterfactuals,* Cambridge University Press, Cambridge.
Makinson, D. (1989), «General Theory of Cumulative Inference», en *Second Int. Workshop in Non Monotonic Reasoning,* Springer Verlag.
Makinson, D. y Gardenfors, P. (1990), «Relations between the Logic of Theory Change and Nonmonotonic Logic», en Fuhrmann y Morreau (eds.), *The Logic of Theory Change,* Springer Verlag, Lect. Notes in Art. Int., 465, 185-205, Berlin.
Makinson, D. (1991), «General patterns in nonmonotonic reasoning», en *Handbook of Logic in AI and Logic Programming,* II, OUP.
McCarthy, J. y Hayes, P. (1969), «Some Philosophical Problems from the Standpoint of Artificial Intelligence», en Meltzer y Michie (eds.), *Machine Intelligence, 4,* 463-502, Edinburgh Univ. Press, Edinburgh.
McCarthy, J. (1980), «Circumscription - a form of non-monotonic reasoning»: *Journal of Artificial Intelligence, 13.*
McCarthy, J. (1986), «Applications of Circumscription to Formalizing Common-Sense Reasoning»: *Artificial Intelligence, 28* (1), 89-116.
McDermott, D. y Doyle, J. (1980), «Non Monotonic Logic I»: *Artificial Intelligence, 13.*
Minsky, M. (1975), «A Framework for Representing Knowledge», en *The Psychology of Computer Vision,* McGraw-Hill.
Moore, R. (1985), «Semantical considerations on Nonmonotonic Logic»: *Artificial Intelligence, 25* (1), 75-94.
Newell, A. y Simon, H. (1976), «Computer Science as Empirical Inquiry: Symbols and Search», en *Communications of ACM,* marzo de 1976.
Newell, A. (1981), «The Knowledge Level»: *Artificial Intelligence Magazine,* verano de 1981.
Nilsson, N. (1991), «Logic and Artificial Intelligence»: *Artificial Intelligence, 47,* 31-56.
Pequeno, T. (1990), «A Logic for Inconsistent Nonmonotonic Reasoning»: *Tech. Report 90/6,* Dept. of Computing Sc. Imperial College, London.
Poole, D. (1988), «A Logical Framework for Default Reasoning»: *Artificial Intelligence, 36,* 27-47.
Reiter, R. (1980), «A Logic for Default Reasoning»: *Artificial Intelligence, 13,* 81-132.
Reiter, R. (1987), «Non Monotonic Reasoning»: *Annual Reviews in Computer Science,* vol. 2.
Ross, W. D. (1927), *The Right and The Good,* Clarendon Press, Oxford.
Shoham, Y. (1987), «A Semantical Approach to Non Monotonic Logics», en *Proceedings Logics in Computer Science,* Ithaca, New York.
Shoham, Y. (1988), *Reasoning about Change,* Cambridge University Press, Cambridge.
Tarski, A. (1956), *Logic, Semantic, Metamathematics. Papers from 1923 to 1938,* Clarendon Press, Oxford.
Touretzky, D. (1986), *The Mathematics of Inheritance Systems: Research Notes in Artificial Intelligence,* Pitman, London.
Wright G. H., von (1971), «A new system of Deontic Logic», en R. Hilpinen (ed.), *Deontic Logic: Introductory and Systematic Readings,* D. Reidel, Dordrecht.

LÓGICA PARACONSISTENTE

Newton C. A. da Costa y Renato A. Lewin

I. INTRODUCCIÓN

Varias son las razones que originaron el surgimiento de la lógica paraconsistente. Mencionaremos las siguientes:

1) En la teoría de conjuntos, tratada en forma intuitiva y no axiomática por Cantor, su creador, y por los matemáticos de fines del siglo pasado, existe un principio básico denominado el axioma de separación (o de comprensión).

Toda propiedad determina un conjunto, aquél formado por los objetos que poseen tal propiedad. Este postulado parece sensato y figuraba, de modo implícito, en el sistema lógico-formal de Frege. Como sabemos, usando leyes y reglas lógicas muy simples, se puede probar que este postulado conduce a contradicción (la paradoja de Russell).

Después del descubrimiento de la paradoja de Russell a comienzos del siglo XX, se hizo necesario axiomatizar la teoría intuitiva de conjuntos, restringiendo el principio de separación para evitar las paradojas. El camino seguido fue el de introducir restricciones al referido principio manteniéndose la lógica clásica (esencialmente el cálculo de predicados de primer orden, con o sin identidad) como lógica subyacente de las teorías de conjuntos obtenidas. Aparecen así las teorías de Zermelo-Fraenkel, de Von Neumann-Bernays-Gödel, de Kelley-Morse, de Quine (*NF* y *ML*), etc.

Surge entonces la pregunta: tanto el principio de separación como la lógica elemental clásica parecen ser plausibles; ¿por qué entonces modificar el primero y conservar la segunda? ¿No sería posible mantener el principio y modificar la lógica tradicional? Evidentemente, en caso de que se quiera proceder de esta manera, la lógica resultante, debe «aceptar» contradicciones ya que el principio en cuestión nos lleva en forma

natural a inconsistencias. En realidad, la lógica clásica (y muchas otras como la intuicionista) es tal que si basamos en ella una teoría en la que podemos derivar una contradicción (una proposición y su negación), entonces la teoría es trivial, en ella se puede «demostrar cualquier cosa». Luego, si queremos desarrollar teorías de conjuntos en las cuales el principio de separación esté sujeto a restricciones más débiles que aquellas de las teorías de conjuntos usuales, o que no esté sujeto a ninguna restricción, debemos emplear lógicas tales que puedan servir de base a teorías inconsistentes (contradictorias), pero no triviales. Estas lógicas se denominan *paraconsistentes*.

Hoy en día se sabe que se puede construir numerosas teorías de conjuntos, inconsistentes no triviales (ver Arruda, 1964, 1970a y 1970b; Arruda y da Costa, 1970, da Costa 1964d, 1965 y 1986).

En general, tales teorías contienen al conjunto de Russell, formado por los conjuntos que no pertenecen a sí mismos y solamente por ellos. Tal conjunto pertenece y no pertenece a sí mismo. En cierto sentido son teorías más fuertes que las teorías de conjuntos usuales ya que, además de los conjuntos normales, clásicos, poseen conjuntos «inconsistentes», como el de Russell. Puede demostrarse también que muchas de esas teorías son no triviales si y solamente si ciertas teorías clásicas, como la de Zermelo-Fraenkel, son consistentes (ver da Costa, 1986).

2) Se sabe que para ciertos cultores de la dialéctica (como en el caso de Hegel, según algunos de sus intérpretes), esa disciplina encierra contradicciones. Por eso, autores como Popper argumentan que la dialéctica es lógicamente imposible: en efecto, la lógica clásica no puede ser la lógica subyacente a la dialéctica, pues si eso ocurriese, sería trivial; luego, como en la época en que Popper estudió los fundamentos de la dialéctica no se conocía la lógica paraconsistente (algunos pensaban incluso que tal lógica no podría existir, como el mismo Popper), esto probaba lógicamente la imposibilidad de la dialéctica. Ahora bien, la lógica paraconsistente, por si sola, no la justifica pero evidencia que las críticas a su estructura lógica, como las señaladas, son infundadas. La lógica de la dialéctica, en conformidad con algunas de sus interpretaciones, tiene que ser paraconsistente (ver da Costa y Wolf, 1980).

3) El filósofo austríaco Meinong desarrolló una teoría de los objetos, en la cual objetos como el círculo cuadrado son legítimos; ver Meinong (1907). No podemos aquí entrar en detalles sobre esa teoría, sólo nos limitaremos a señalar que Bertrand Russell la criticó especialmente por conducir a contradicciones, al infringir la ley de contradicción. Una manera de superar tal dificultad sería el utilizar una lógica paraconsistente como lógica básica de la teoría de Meinong.

4) La lógica paraconsistente nació también del deseo de esclarecer mejor determinadas cuestiones lógicas. Por ejemplo, ¿qué es la negación? La negación clásica posee ciertas propiedades, pero hay varias negaciones paraconsistentes que poseen propiedades análogas y que pueden, por lo tanto, también ser tenidas por negaciones. Aquí sucede algo similar

a lo que sucede con el concepto de recta, que puede ser tanto la recta de la geometría euclidiana como de las no-euclidianas.

5) Existen teorías paraconsistentes de la verdad que extienden la teoría tarskiana. Esto significa que hay semánticas alternativas de la semántica clásica, así como hay geometrías distintas de la geometría euclidiana, mereciendo todas ser consideradas como geometrías. El deseo de saber si había semánticas paraconsistentes fue otro de los motivos de la creación de la lógica paraconsistente.

6) La manipulación sensata de sistemas inconsistentes de manejo de información, por ejemplo como se da hoy en inteligencia artificial, fue otra razón para elaborar una lógica paraconsistente. Una situación similar ocurre cuando se trata de sistematizar logicamente códigos éticos o jurídicos, que en general son inconsistentes, sin intentar desfigurarlos haciéndolos consistentes. Hacer esto o bien es imposible en la práctica, o bien se transforma tales códigos en otra cosa: ya no estamos hablando de lo mismo. Por eso, el tratamiento paraconsistente de los códigos éticos o jurídicos constituye una posible solución aunque existen autores que proponen otras alternativas. En tanto, en inteligencia artificial parece no haber indicaciones de una solución alternativa a la paraconsistente.

Otras razones para la introducción de la lógica paraconsistente resultarán evidentes a medida que avance nuestra exposición.

Podemos pasar ahora a una presentación más rigurosa de la lógica paraconsistente.

Una teoría deductiva T se caracteriza por su lenguaje L, por su lógica $\mathcal{L}$ y por sus principios específicos (axiomas y postulados). Supongamos que L contiene un símbolo para la negación (por ejemplo, $\neg$). Se dice entonces que T es *trivial* si todos sus enunciados (fórmulas, muchas veces sólo nos interesan las fórmulas cerradas u oraciones) son teoremas. En caso contrario se dice que T es no trivial. La teoría T se dice *inconsistente* (o contradictoria) si ella contiene, al menos, dos teoremas de la forma α y $\neg\alpha$, uno de los cuales es la negación del otro. En caso contrario T se dice consistente.

Una lógica se dice *paraconsistente* si puede ser la lógica de teorías inconsistentes pero no triviales. Una tal teoría se denomina teoría paraconsistente. Luego una lógica es paraconsistente si puede ser la lógica subyacente de teorías paraconsistentes. Evidentemente la mayoría de las lógicas usuales, como la clásica, no son paraconsistentes. Por otro lado en una lógica paraconsistente se puede, a veces, basar teorías consistentes; además, el que una lógica $\mathcal{L}$ sea paraconsistente no implica que una teoría basada en ella no pueda ser trivial.

Dada un lógica $\mathcal{L}$, puede ocurrir que el lenguaje L contenga más de una negación, digamos $\neg$ y $\sim$. En este caso, $\mathcal{L}$ puede ser paraconsistente con respecto a una de estas negaciones, por ejemplo $\neg$, pero no con respecto a la otra negación $\sim$. Situaciones como ésa (y otras aún más complicadas) han sido consideradas, pero para los objetivos de este trabajo, no vamos a estudiarlas.

El principio de contradicción tiene varias formulaciones que no son equivalentes entre sí. Para nosotros, las dos siguientes son importantes:

I. Dadas dos proposiciones α y $\neg\alpha$, una de las cuales es la negación de la otra, una de ellas es falsa.
II. La proposición $\neg(\alpha\wedge\neg\alpha)$ es verdadera, donde α es una proposición cualquiera, $\neg$ es el símbolo de negación y $\wedge$ representa el conectivo de conjunción.

En una lógica paraconsistente $\mathcal{L}$, la formulación I del principio de contradicción no puede ser válida. En efecto, si $\mathcal{L}$ es paraconsistente existe al menos una teoría *T*, basada en $\mathcal{L}$, que tiene como teoremas proposiciones de la forma α y $\neg\alpha$; entonces α y $\neg\alpha$ deben ser ambas verdaderas en *T* y el principio es violado. En tanto, en la formulación II, el principio puede valer en una lógica paraconsistente.

Así, hay leyes y reglas que no pueden ser válidas en una lógica paraconsistente. Si $\rightarrow$ y $\vee$ representan respectivamente la implicación (que satisface la regla de Modus Ponens) y la disyunción, mencionaremos aquí algunos principios que no son válidos: $\alpha\rightarrow(\neg\alpha\rightarrow\beta)$, $\neg\alpha\rightarrow(\alpha\rightarrow\beta)$, $(\alpha\wedge\neg\alpha)\rightarrow\beta$, $((\alpha\vee\beta)\wedge\neg\alpha)\rightarrow\beta$.

En relación a la lógica paraconsistente en nuestra opinión se puede asumir dos posiciones:

1) La lógica paraconsistente es considerada rival de la lógica clásica, destinada a sustituirla en todos o algunos campos del saber.

2) Ella puede considerarse un complemento de la lógica clásica; esta última sería la lógica básica aplicable en principio a todas las circunstancias. La lógica paraconsistente constituiría una especie de formalismo lógico-matemático que en ciertas situaciones, por ejemplo en inteligencia artificial, debería ser usado por motivos de conveniencia práctica aunque la lógica clásica continúe operando.

Decidir, en caso de que sea posible, cuál de esas posiciones es la correcta, lo que constituye un problema de lógica filosófica, es importante pero no será tratado aquí. El desarrollo de la lógica paraconsistente como disciplina científica importante no depende de su solución. De cualquier forma el nacimiento de la lógica paraconsistente produjo un cambio en el paradigma en el campo de la lógica, de la ciencia y de la filosofía, especialmente si la consideramos como rival de la lógica clásica.

Algunas lógicas paraconsistentes difieren mucho de la clásica (por ejemplo la lógica de la paradoja de Priest, 1979). Sin embargo, hay sistemas paraconsistentes, que aunque difieren del clásico, lo contienen como una parte que se aplica en ciertos casos, que dan origen a una matemática paraconsistente más fuerte que la clásica y que contienen a ésta propiamente y significativamente (ver da Costa, 1974b).

Para terminar esta introducción, creemos conveniente hacer algunos comentarios sobre una lógica que es la «dual» de la paraconsistente: la lógica paracompleta.

El principio del tercero excluido es susceptible de diversas formulaciones alternativas, pero no equivalentes. Aquí nos interesarán dos:

I. Dadas dos proposiciones contradictorias α y $\neg\alpha$, una de ellas es verdadera.
II. La proposición $\alpha \vee \neg\alpha$ es verdadera.

Esencialmente, una lógica paracompleta viola al menos una de estas formulaciones del «tertium non datur». La lógica intuicionista de Brouwer y Heyting es paracompleta; lo mismo ocurre con varias lógicas polivalentes.

Normalmente toda lógica paraconsistente tiene una dual paracompleta y recíprocamente. Existen lógicas paracompletas que no son paraconsistentes y viceversa. Cuando una lógica es simultáneamente paraconsistente y paracompleta se le llama no-alética. Lógicas de esta última categoría han sido empleadas en los campos de la ética y del derecho.

II. ASPECTOS HISTÓRICOS DE LA LÓGICA PARACONSISTENTE

La paraconsistencia es el conjunto de temas históricos, filosóficos y científicos relacionados con el concepto de contradicción; engloba temas como la filosofía de Heráclito, Hegel y Marx, la naturaleza y los límites de la dialéctica, contradicción y realidad y el significado de una ontología paraconsistente (esto es, basada en una lógica paraconsistente). Actualmente la lógica paraconsistente constituye un asunto técnico y científico bien definido, al igual que, por ejemplo, la lógica tradicional, y no se confunde con la paraconsistencia en general.

No nos ocuparemos aquí de la paraconsistencia en general sino sólo de la historia de la lógica paraconsistente como disciplina científica. Además sólo trataremos los aspectos que creemos más importantes de esa historia. Para tener una idea más acabada del tema el lector puede consultar, por ejemplo, Arruda (1980 y 1984), da Costa y Marconi (1986), D'Ottaviano (1990) y Priest, Routley y Norman (1989). Una introducción elemental a la lógica paraconsistente, con referencias históricas es Grana (1983); Marconi (1979) también contiene abundantes observaciones de índole histórica. La lógica relevante, es claro, está íntimamente ligada a la lógica paraconsistente; por lo tanto cabe mencionar también la obra de Routley *et al.* (1983), la cual contiene numerosas referencias históricas pertinentes.

Dos precursores de la lógica paraconsistente fueron el lógico polaco J. Lukasiewicz y el filósofo ruso N. A. Vasilev. Ambos en 1910, en forma totalmente independiente, estudiaron la posibilidad de una lógica paraconsistente. El primero, en un artículo bien conocido (1910, traducción inglesa, Lukasiewicz, 1971), discurrió sobre una lógica donde no fuera válida la ley de contradicción en alguna de sus formas. Ahora bien, en aquella época no es extraño que él sólo considerase la lógica tradicional

en su formulación aristotélica; no obstante, el hecho es que él habla de la derogación de la ley de contradicción, lo que implica el aceptar tener proposiciones contradictorias verdaderas (esto es, pares de fórmulas contradictorias α y $\neg\alpha$, simultáneamente verdaderas). Sin embargo, él no elabora ningún sistema explícito de lógica paraconsistente.

Vasilev, por su lado, modifica la lógica en su presentación aristotélica, construyendo lógicas imaginarias, las cuales no eliminaban la existencia de contradicciones verdaderas (ver Vasilev, 1925). Sin embargo, el lógico ruso no desarrolló sus sistemas dentro de los patrones de rigor y amplitud de la lógica contemporánea, permaneciendo prisionero de la concepción aristotélica de la lógica. La obra de Vasilev ha sido muy estudiada últimamente, especialmente por autores soviéticos como W. Smirnov y V. A. Bazhanov (Bazhanov, 1989, en una obra muy buena). La lógica brasileña A. I. Arruda dedicó varios artículos a la obra de Vasilev (ver, por ejemplo, Arruda, 1977 y Arruda, 1984).

El primer autor en formular un cálculo proposicional paraconsistente fue el lógico polaco S. Jaśkowski, según él mismo afirma, a sugerencia de Lukasiewicz, en 1948. Él llamó a su cálculo lógico discusivo o discursivo, ya que una de las motivaciones para construir un sistema era la siguiente: si queremos reunir en una única teoría todas las afirmaciones hechas en una discusión, como los términos usados no son empleados siempre con el mismo sentido, muchas veces sucederá que la teoría considerada contendrá proposiciones contradictorias, defendidas por los diferentes participantes de la discusión o por el mismo participante en momentos distintos. Así, si debe evitarse la trivialización, la lógica de una teoría como ésa deber ser paraconsistente.

Jaśkowski no axiomatizó su cálculo proposicional. Tan sólo lo definió por intermedio de una interpretación en el sistema modal S5 de Lewis. No fue sino hasta varios años después que su sistema fue axiomatizado en da Costa y Dubikajtis (1968 y 1977) y extendido a un cálculo de predicados de primer orden y de orden superior, ver también da Costa (1975), Kotas (1975) y Kotas y da Costa (1977). Hoy en día la lógica discusiva está bien desarrollada y ha encontrado numerosas aplicaciones; por ejemplo, es la lógica subyacente de una nueva conceptualización de verdad pragmática, como se demuestra en da Costa y Chuaqui (en prensa). Se puede constatar fácilmente que la lógica discusiva también puede ser interpretada como una lógica de la vaguedad.

En el presente estado de evolución de la lógica, para tener una lógica o sistema lógico propiamente tal, es necesario que esté desarrollado al menos un cálculo de predicados de primer orden con identidad; en otras palabras, se hace imprescindible que se sepa operar con los conectivos logícos, pero además, con cuantificadores y la identidad.

Si adoptamos ese punto de vista, podemos decir que el verdadero creador de la lógica paraconsistente fue N. C. A. da Costa. En efecto, en las décadas de los 50 y de los 60, en Brasil e independientemente de los trabajos de Lukasiewicz y de Vasilev, cuyas investigaciones sobre para-

consistencia, así como los trabajos de Jaśkowski, quedaron prácticamente olvidadas, N. C. A. da Costa construyó jerarquías infinitas de cálculos lógicos paraconsistentes, cálculos proposicionales, cálculos de predicados de primer orden, con y sin identidad, cálculos de descriptores y teorías de conjuntos paraconsistentes, ver da Costa (1958-1974). Con sus discípulos brasileños desarrolló enormemente tal lógica, desarrolló lógicas paraconsistentes de distinta naturaleza, por ejemplo sistemas relacionados con la vaguedad en Arruda y Alves (1979a y 1979b); lógica polivalente en D'Ottaviano (1982, 1985a y 1985b); lógica relevante en Arruda y da Costa (1966 y 1984); lógicas deónticas y modales en da Costa y Carnielli (1986), Puga, da Costa y Carnielli (en prensa) y Puga (1985). Numerosos otros desarrollos de la lógica paraconsistente efectuados por miembros de la escuela brasilera de lógica, tales como A. Loparic, C. A. A. P. Abar, E. H. Alves, A. I. Arruda, W. Carnielli, L. H. dos Santos, I. M. L. D'Ottaviano, L. de Moraes, W. da Silva, J. Abe, D. Krause, L. P. de Alcántara y A. M. Sette, son descritos en D'Ottaviano (1990) y da Costa y Marconi (1989).

La lógica paraconsistente dio origen a varios desarrollos técnicos que tienen un significado que trasciende el campo de la paraconsistencia. Aunque no podemos entrar en detalles, creemos conveniente señalar los siguientes:

1) La algebrización de ciertas lógicas paraconsistentes no puede hacerse mediante las técnicas tradicionales de la lógica algebraica, a través de las denominadas álgebras de Lindenbaum. Luego fue preciso que se elaborasen nuevas técnicas como se hizo por ejemplo en da Costa, Sette (1969).

2) Se creó semánticas para las lógicas paraconsistentes; un método para ello es el método de las valuaciones (ver Grana, 1990b). Se verificó entonces que la teoría de la verdad de Tarski, como dijimos anteriormente, puede ser ampliada al caso en el que hay contradicciones «verdaderas». También Routley y Meyer construyeron semánticas para la lógica relevante que se aplican al caso de la paraconsistencia (ver Routley y Meyer, 1976, y Routley, 1979).

3) Se está desarrollando una matemática paraconsistente. Para citar dos ejemplos, ya se estudió una geometría afín paraconsistente en da Costa (1989), y en Mortensen (1990) se formula una versión paraconsistente del cálculo diferencial. Notemos, para terminar esta referencia a la matemática, que ésta está íntimamente relacionada con la matemática «fuzzy».

La lógica paraconsistente se ha convertido en una de las disciplinas más cultivadas en el mundo, en gran medida, por sus aplicaciones a las ciencias de la computación, en especial, a la inteligencia artificial.

Describiremos ahora superficialmente lo que se ha hecho en el campo de la lógica paraconsistente en algunos centros fuera de Brasil.

En Estados Unidos, C. Pinter (1980) investigó un interesante sistema lógico, emparentado con la lógica discusiva de Jaśkowski que se deno-

mina lógica de la ambigüedad inherente, la cual tiene aplicaciones en inteligencia artificial y en lingüística. R. L. Epstein (1990) se ocupó de la paraconsistencia dentro del campo de una visión propia de la lógica. R. C. Wolf trató cuestiones filosóficas relacionadas con las lógicas paraconsistentes y relevantes (da Costa y Wolf, 1980 y 1985). N. Belnap (1977) trabajando en fundamentos de computación formuló una lógica tetravalente que es paraconsistente. Otros lógicos y filósofos como J. M. Dunn y S. French han trabajado en tópicos cercanos a la paraconsistencia. Merece destacarse el artículo pionero de Nelson (1959), y el libro de N. Rescher y R. Brandon (1964).

En Chile, R. Chuaqui trató la lógica de la verdad pragmática, en una de sus posibles interpretaciones, evidenciando con da Costa que ella es una lógica discusiva (ver da Costa y Chuaqui, en prensa). Otros lógicos chilenos han estudiado fundamentalmente aspectos algebraicos de la lógica paraconsistente; aplicando los conceptos desarrollados por Blok y Pigozzi han establecido la algebrizabilidad de algunos sistemas como el sistema P_1 de Sette (1973). Asimismo, han demostrado la no algebrizabilidad de otros simplificando demostraciones anteriores (ver Lewin, Mikenberg y Schwarze, 1990; 1991). También han estudiado las álgebras que se obtienen en el proceso de algebrización (ver Lewin, Mikenberg y Schwarze, 1994).

Dentro de lo realizado en Argentina, recordemos el trabajo de lógica algebraica paraconsistente de M. Fidel (1977) y las investigaciones de A. R. Raggio en la formulación estilo Gentzen de ciertos sistemas paraconsistentes en Raggio (1968). El lógico argentino F. Asenjo, hoy radicado en Estados Unidos, construyó una lógica antinómica y trató los fundamentos de la teoría de conjuntos, obteniendo, con J. Tamburino, una teoría de conjuntos paraconsistente extremadamente fuerte (ver Asenjo, 1966 y Asenjo y Tamburino, 1975). Asenjo fue uno de los pioneros de la lógica paraconsistente, sus estudios iniciales se hicieron independientemente de los de da Costa y otros.

En Uruguay hay un pequeño grupo de lógicos que se interesan sobre todo por las aplicaciones a la filosofía de la lógica paraconsistente. C. E. Caorsi ha intentado aplicar ciertos sistemas paraconsistentes a los fundamentos del psicoanálisis.

El lógico F. Miró Quezada fue quien acuñó en 1976 la palabra «paraconsistencia» para designar la nueva lógica; también se deben a él los términos «paracompleto» y «no-alético». Miró Quezada se ha ocupado, principalmente, de la filosofía de la lógica paraconsistente (ver, por ejemplo, su artículo en Priest, Routley y Norman, 1989).

En Israel, A. Avron (1990a y 1990b) desarrolló una versión nueva tanto de la paraconsistencia como de la relevancia. Sus resultados son significativos desde el punto de vista teórico y tienen aplicaciones a la informática.

En Australia se encuentra uno de los grupos más fuertes de lógicos que se dedican a temas paraconsistentes. No es posible analizar aquí todo

lo que se ha hecho en Australia, nos limitaremos pues a hacer algunas indicaciones. Routley y Meyer estudiaron una forma de lógica paraconsistente y relevante que bautizaron lógica dialéctica (ver Routley y Meyer, 1976, importante sobre todo desde el punto de vista filosófico). Ellos desarrollaron una semántica para la lógica relevante que también tiene aplicaciones en el campo de la paraconsistencia. Routley trató temas como lógica deóntica paraconsistente, teoría de Meinong, teoría de la decisión, cálculo de probabilidades, etc. G. Priest, como ya observamos, creó un sistema paraconsistente, que ha intentado aplicar prácticamente a todos los problemas más importantes de la filosofía, la ciencia y la técnica. C. Mortensen ya estructuró un cálculo diferencial paraconsistente y está desarrollando una matemática paraconsistente (la que incluye el álgebra, el análisis, la geometría y la mecánica). Otro lógico importante radicado en Australia es M. Bunder con una vasta contribución técnica (ver Bunder, 1974-1989).

En la Unión Soviética, la lógica paraconsistente encontró un campo fértil para prosperar. Entre los especialistas soviéticos que desarrollaron nuevos sistemas, la historia o aplicaciones de la paraconsistencia mencionaremos a W. Smirnov, V. A. Bazhanov, A. S. Karpenko e I. S. Narski. El artículo de Karpenko (1984) es una muestra significativa de lo realizado en la Unión Soviética.

La contribución polaca a la paraconsistencia es fundamental. Fuera de Jaśkowski, mencionaremos a L. Dubikajtis y sus discípulos, quienes se dedicaron principalmente a la axiomatización de sistemas paraconsistentes; también J. Kotas, quien elaboró simultáneamente sistemas polivalentes y paraconsistentes y profundizó una investigación de la lógica discusiva.

En Italia, la obra de D. Marconi, N. Grana, S. Coradeschi, P. Bottura y M. L. Dalla Chiara, entre otros, merecen mención. Dalla Chiara empleó técnicas paraconsistentes en los fundamentos de la física, en particular, de la mecánica cuántica. N. Grana (1983 y 1990a) escribió la primera introducción elemental a la lógica paraconsistente y estudió la lógica deóntica paraconsistente. D. Marconi, entre otras contribuciones, trató la lógica de Hegel bajo el prisma de la lógica paraconsistente.

El filósofo español L. Peña (1979 y 1980) construyó una lógica paraconsistente de índole *fuzzy* basado en un profundo análisis filosófico. Fuera de eso ha aplicado sus ideas lógicas a los más variados tópicos, como por ejemplo, la dialéctica.

En Francia (M. Guillaume, J. V. Béziau,...), en Bulgaria (H. Smolenov, S. Petrov,...) y otros países también se ha cultivado la lógica paraconsistente.

Por motivos obvios una exposición histórica como la anterior no puede ser ni completa ni equilibrada, sin embargo, creemos que las referencias bibliográficas ayudarán al lector a completarla.

III. ALGUNOS ASPECTOS TÉCNICOS

Describiremos en esta sección algunos sistemas deductivos paraconsistentes, aspectos semánticos de éstos y por último algunos aspectos algebraicos.

1. *El sistema de Jaśkowski*

Como dijimos Jaśkowski no introdujo un sistema paraconsistente propiamente tal sino que hizo una interpretación de sus conectivos discusivos en la lógica modal **S5** de Lewis. Su conjunción, implicación y equivalencia discusiva se definía como sigue

$$p \wedge_d q := p \wedge \Diamond q,$$
$$p \rightarrow_d q := \Diamond p \rightarrow q$$
$$p \leftrightarrow_d q := (\Diamond p \rightarrow q) \wedge (\Diamond q \rightarrow p),$$

donde $\wedge$, $\rightarrow$ y $\Diamond$ son los funtores de **S5**. El operador $\Diamond$ se puede leer como «alguien afirma que».

En da Costa y Dubikajtis (1968), se introduce la siguiente axiomatización, el sistema **J**, para este cálculo y se inicia su estudio semántico:

$$J_1 : \Box((A \rightarrow B) \rightarrow ((B \rightarrow C) \rightarrow (A \rightarrow C)))$$
$$J_2 : \Box((\neg A \rightarrow A) \rightarrow A)$$
$$J_3 : \Box(B \rightarrow (A \vee B))$$
$$J_4 : \Box(\Box(A \rightarrow B) \rightarrow \Box(\Box A \rightarrow \Box B)$$
$$J_5 : \Box((A \rightarrow C) \rightarrow ((B \rightarrow C) \rightarrow ((A \vee B) \rightarrow C)))$$
$$J_6 : \Box(A \rightarrow (\neg A \rightarrow B))$$
$$J_7 : \Box(A \rightarrow (A \vee B))$$
$$J_8 : \Box(\Box A \rightarrow A)$$
$$J_9 : \Box(A \rightarrow \Box \Diamond A)$$
$$R_1 : \frac{A, \Box(A \rightarrow B)}{B}$$
$$R_2 : \frac{\Diamond A}{A}$$

Allí prueban que el sistema **J** es equivalente a la lógica de Jaśkowski. En un trabajo posterior (da Costa, Dubikajtis, 1977), los autores introducen sistemas de lógica discusiva de orden superior basados en $\mathbf{S5}_\omega$ y definen una semántica que extiende a los modelos de Kripke para lógica modal.

Resulta claro que a cualquier lógica modal puede asociarse la correspondiente lógica de Jaśkowski. Éstas fueron estudiadas en Kotas y da Costa (1977), siendo algunas de ellas interesantes.

2. *La jerarquía* C_n, $1 \leq n \leq \omega$, de da Costa

El cálculo C_1 tiene como símbolos primitivos variables proposicionales $\neg$, $\vee$, $\wedge$, $\rightarrow$ y paréntesis. Los axiomas y reglas de C_1 son los siguientes:

$$
\begin{aligned}
A_1 &: A\to(B\to A)\\
A_2 &: (A\to B)\to((A\to(B\to C))\to(A\to C))\\
A_3 &: A\to(B\to A\wedge B)\\
A_4 &: A\wedge B\to A\\
A_5 &: A\wedge B\to B\\
A_6 &: (A\to C)\to((B\to C)\to((A\vee B)\to C))\\
A_7 &: A\to A\vee B\\
A_8 &: B\to A\vee B\\
A_9 &: \neg\neg A\to A\\
A_{10} &: A\vee\neg A\\
A_{11} &: B^{\circ}\to((A\to B)\to((A\to\neg B)\to\neg A))\\
A_{12} &: A^{\circ}\wedge B^{\circ}\to(A\to B)^{\circ}\wedge(A\wedge B)^{\circ}\wedge(A\vee B)^{\circ}\\
MP &: \frac{A, A\to B}{B}
\end{aligned}
$$

donde $A^{\circ} := \neg(A\wedge\neg A)$. Intuitivamente, A° significa que A se comporta «bien», es decir, no es contradictoria. Esto se fundamenta en el axioma A_{11}, el que no dice otra cosa que el principio de reducción al absurdo se puede aplicar siempre que la oración que se «corta» no sea contradictoria. El axioma A_{12} nos dice que el buen comportamiento se extiende a las oraciones complejas. Para justificación intuitiva de estos axiomas ver da Costa y Carnielli (1986).

Para $1\leq n<\omega$ definimos

$$
\begin{aligned}
A^{n} &:= A^{\circ\cdots\circ},\ n \text{ veces y}\\
A^{(n)} &:= A^{\circ}\wedge A^{\circ\circ}\wedge\ldots\wedge A^{n}.
\end{aligned}
$$

Los cálculos C_n, $1<n<\omega$, se obtienen reemplazando los axiomas A_{11} y A_{12}, respectivamente, por

$$
\begin{aligned}
A_{11}^{n} &: B^{(n)}\to((A\to B)\to((A\to\neg B)\to\neg A))\\
A_{12}^{n} &: A^{(n)}\wedge B^{(n)}\to(A\to B)^{(n)}\wedge(A\wedge B)^{(n)}\wedge(A\vee B)^{(n)}.
\end{aligned}
$$

Finalmente C_{ω} está definido por los axiomas A_1—A_{10} y la regla MP. En los cálculos C_n, $n<\omega$, se define la negación fuerte

$$\neg^{*}A := \neg A\wedge A^{(n)}$$

No es difícil verificar que $\neg^{*}$ tiene todas las propiedades de la negación clásica.

Algunos teoremas importantes que pueden encontrarse en da Costa (1958-1974) son los siguientes:

Teorema: *Todas las reglas y esquemas válidos del cálculo proposicional positivo clásico son válidos en C_n, $i\leq n<\omega$.*

Teorema: *Si Γ es un conjunto de oraciones y $A_1,\ldots, A_m$ son las componentes primas de las fórmulas de $\Gamma\cup\{A\}$, entonces $\Gamma\vdash_{C_0} A$ si y sólo si $\Gamma, A_m^{(n)}{}_{C_0} A$, para $1\leq n<\omega$, donde C_0 es el cálculo proposicional clásico.*

En virtud de este teorema, todos los teoremas de C_0 son válidos en C_n, $1 \leq n < \omega$, para las oraciones «buenas».

Un sistema no trivial S es finitamente trivializable si existe una fórmula F tal que el sistema obtenido al agregar F a S, como nuevo axioma, es trivial.

Teorema: *Los sistemas C_n, $1 \leq n \leq \omega$, son no triviales. Los C_n, $1 \leq n < \omega$ son finitamente trivializables, pero C_ω no lo es.*

Otro resultado interesante que aparece en Alves (1976) es el siguiente:

Teorema: *Los axiomas de C_n, $1 \leq n < \omega$, son independientes. Cada sistema es estrictamente más fuerte que los que lo siguen.*

En Urbas (1989) se hace notar que los cálculos C_n, $1 \leq n < \omega$, no tienen la propiedad de sustitución de equivalentes; también se demuestra allí que si agregamos reglas para remediarlo, por ejemplo $RC : \dfrac{C \to D}{\neg D \to \neg C}$

todos los sistemas excepto C_ω, colapsan a C_0. Esta debilidad está en la base de la no algebrizabilidad de estos sistemas, como veremos más adelante.

3. *Las jerarquías* C_n^* y $C_n^=$, $1 \leq n < \omega$

Da Costa extendió su jerarquía al cálculo de predicados de primer orden y de primer orden con identidad obteniendo los sistemas C_n^* y $C_n^=$, $1 \leq n < \omega$, respectivamente.

Los axiomas para C_n^*, $1 \leq n < \omega$ son los de C_n, $1 \leq n < \omega$, más los siguientes:

A_1^*: Si A y B son fórmulas congruentes (ver Kleene, 1952), o una se obtiene de la otra eliminando cuantificadores vacíos, entonces $A \leftrightarrow B$ es un teorema.

A_2^*: $\forall x(A(x))^{(n)} \to (\forall x A(x))^{(n)}$
A_3^*: $\forall x(A(x))^{(n)} \to (\exists x A(x))^{(n)}$

Los postulados de C_ω^* son los de C_ω más los del cálculo de predicado clásico y A_1^*.

Los postulados de $C_n^=$, $1 \leq n < \omega$, son los de C_n^* más los axiomas usuales para la identidad.

C_0^* obtenido de C_1^* agregando el nuevo postulado $\neg(A \wedge \neg A)$, es el cálculo de predicados clásico.

A modo de ilustración, las siguientes oraciones no son válidas en C_1^*:

$$\neg \exists x \neg A(x) \leftrightarrow \forall x A(x)$$
$$\neg \forall x \neg A(x) \leftrightarrow \exists x A(x).$$

Teorema: *Los cálculos C_n^* y $C_n^=$, $1 \leq n < \omega$, son indecidibles.*

Por último, podemos decir que los cálculos C_n^*, $1 \leq n < \omega$, son extensiones conservadoras de los respectivos cálculos C_n. Lo mismo ocurre para $C_n^=$ y C_n^*, $1 \leq n < \omega$.

4. *Semántica*

En da Costa y Alves (1977) se presenta una semántica estilo Henkin sobre la base de una generalización del concepto clásico de valuación. Con ellos demuestran que los sistemas de da Costa son completos.

Una valuación para C_n, $1 \leq n < \omega$, es una función $v: F \rightarrow \{0,1\}$, donde F es el conjunto de las fórmulas de C_n, tal que

i) Si $v(A) = 0$, entonces $v(\neg A) = 1$,
ii) Si $v(\neg \neg A) = 1$, entonces $v(A) = 1$,
iii) Si $v(B^{(n)}) = v(A \rightarrow B) = v(A \rightarrow \neg B) = 1$, entonces $v(A) = 0$,
iv) $v(A \rightarrow B) = 1$ si y sólo si $v(A) = 0$ o $v(B) = 1$,
v) $v(A \wedge B) = 1$ si y sólo si $v(A) = 1$ y $v(B) = 1$,
vi) $v(A \vee B) = 1$ si y sólo si $v(A) = 1$ o $v(B) = 1$,
vii) Si $v(A^{(n)}) = v(B^{(n)}) = 1$, entonces $v((A \rightarrow B)^{(n)}) = v((A \wedge B)^{(n)})$ $= v((A \vee B)^{(n)}) = 1$.

Una valuación es un modelo de un conjunto de oraciones Γ si y sólo si para todo $A \in \Gamma$, $v(A) = 1$. El concepto de consecuencia semántica $\Gamma \models A$ se define en la forma habitual.

Podemos enunciar los teoremas demostrados en las obras citadas.

Teorema: *Todo conjunto maximal no trivial de oraciones de C_n, $1 \leq n < \omega$, (consistente o no) tiene modelo.*

Teorema: $\Gamma \vdash_{cn} A$ *si y sólo si* $\Gamma \models A$, $1 \leq n < \omega$.

Teorema: *Los cálculos C_n, $n \leq \omega$ son decidibles.*

Cabe aquí destacar que en Arruda (1975) se demuestra que estos sistemas no son decidibles por matrices finitas. En Arruda y da Costa (1977) se extiende el método de valuaciones a los sistemas C_n^*, $1 \leq n < \omega$. El caso de C_ω^* y $C_\omega^=$ requieren de un tratamiento especial.

5. *Aspectos algebraicos*

La construcción del álgebra de Lindenbaum de un sistema lógico ha producido grandes frutos, sin embargo, no puede aplicarse a la mayoría de los sistemas paraconsistentes. El motivo de esto es que la relación de equivalencia correspondiente no es una congruencia, o bien, no es compatible con el conjunto de los teoremas del sistema. Si bien el concepto de Lindenbaum-algebrizable estaba claro, no resultaba claro si existen otros métodos de algebrizar una lógica. El motivo de esto es que no existía una teoría general de la algebrizabilidad de sistemas deductivos. Creemos que el trabajo reciente de Blok y Pigozzi (1989), provee ese marco.

Se ha intentado algebrizar los cálculos C_n, $1 \leq n \leq \omega$, al menos de tres maneras distintas. Da Costa (1966b), y da Costa y Sette (1969) presentaron una forma de algebrización que refleja algunas propiedades de sus sistemas lógicos.

Carnielli y de Alcántara (1984) presentan en términos conjuntistas una versión algebraica de C_1, a la que llaman álgebras paraconsistentes de conjuntos. Éstas contienen un cuasi-orden que representa la deducibilidad lógica del sistema C_1.

Estas algebrizaciones, si bien reflejan algunas propiedades algebraicas de los sistemas, tienen la dificultad de que no se puede identificar fórmulas intuitivamente equivalentes, luego los modelos son más complicados.

Mortensen (1980) contiene la primera demostración de que C_1 (y por lo tanto C_n, $n<1$) no es Lindenbaum-algebrizable. Allí se prueba que en el álgebra absolutamente libre de fórmulas de C_1, no existe una relación de congruencia no trivial que sea compatible con los teoremas de C^1. Lewin, Mikenberg, Schwarze (1991) contiene una versión simplificada de este resultado.

Usando los métodos desarrollados en Blok y Pigozzi (1989), se ha algebrizado la lógica $\boldsymbol{P}^1$ de Sette (1973) y la lógica $\boldsymbol{J}_3$ de D'Ottaviano y da Costa (1970), el sistema de Batens (1980) y otros. Estos resultados pueden encontrarse en Lewin, Mikenberg y Schwarze (1990), Blok y Pigozzi (1989), y Lewin, Mikenberg y Schwarze (1989).

IV. PRINCIPALES APLICACIONES DE LA LÓGICA PARACONSISTENTE

Dividiremos las aplicaciones más significativas de la lógica paraconsistente en aplicaciones filosóficas, aplicaciones científicas y aplicaciones tecnológicas. De la exposición anterior se entrevé muchas de tales aplicaciones. En esta sección nos limitaremos a mencionar algunas otras que, en nuestra opinión, son de gran importancia.

1. *Aplicaciones filosóficas*

Gracias a la lógica paraconsistente, la lógica dialéctica y la teoría de los objetos de Meinong tienen formulaciones lógicamente inobjetables. También, comprendemos mejor la noción de negación y sus posibles variantes; así también, como lo hicimos notar, existe una teoría de la verdad, similar a la de Tarski, que es paraconsistente. Claramente todo esto conlleva revisiones de algunas tesis filosóficas las que, para ser justificadas, requieren de nuevas indagaciones lógicas.

Otra aplicación se refiere a la ontología, la disciplina de las características más generales de lo que existe. Si se usa la lógica tradicional como lógica de la ontología, entre los objetos existentes no se encuentran, automáticamente, ciertos objetos «inconsistentes», como por ejemplo el con-

junto de Russell. Sin embargo, cuando recurrimos a una lógica paraconsistente, todo cambia. Como dijimos, hay teorías de conjuntos donde el conjunto de Russell «existe». Luego, una ontología fundada en una lógica paraconsistente puede, en principio, contener objetos contradictorios. Aceptar o no esa tesis implica, obviamente, que se argumente en profundidad y se analice los cimientos tanto de la lógica como de la ontología. En cierto sentido se puede sustentar que mientras más débil sea nuestra lógica, tanto más rica es nuestra ontología.

La propia lógica paraconsistente nos obliga a repensar la misma noción de lógica. ¿Es la lógica paraconsistente una verdadera lógica? Estos problemas, que surgen en relación a todas las logicas no-clásicas, propuestas como rivales de la lógica clásica, asumen un cariz dramático en lo tocante a la lógica paraconsistente ya que ésta deroga el principio de contradicción, generalmente considerado como el más evidente de todos. Lo que la lógica en cuestión parece demostrar es que existe logicidad incluso si este principio es limitado. Sin embargo, todo indica que los sistemas paraconsistentes necesitan de él, al menos en parte, para su desarrollo.

La utilización de la lógica paraconsistente hace repensar muchos problemas filosóficos de extraordinario interés y en eso reside una de las más sobresalientes aplicaciones de la misma en el campo de la filosofía.

2. *Aplicaciones a la ciencia*

En matemática, la lógica paraconsistente ha originado nuevas ideas y métodos. Así, hoy existe una teoría de modelos paraconsistentes que, en su formulación general, se convierte en una semántica para cualquier sistema lógico (ver Grana, 1990b). Las versiones paraconsistentes de la geometría y del cálculo dieron origen a nuevos conceptos y estructuras matemáticas. Creemos que eso no destruye la matemática tradicional sino que la amplía y pone en evidencia sus limitaciones. Hay aquí un progreso efectivo y ampliación de horizontes.

Ya mencionamos que la paraconsistencia entró en el campo de la física. Ciertas formulaciones de la mecánica cuántica involucran nociones paraconsistentes como pusieron en evidencia Dalla Chiara y su escuela.

También se ha intentado aplicaciones a la psicología (ver da Costa y French, 1990).

3. *Aplicaciones tecnológicas*

Sin duda las más significativas y atractivas aplicaciones de la lógica paraconsistente se verifican en el dominio de la informática, en particular, de la inteligencia artificial.

Para manipular informaciones inconsistentes, no existen sistemas lógicos apropiados y como dejamos claro anteriormente, éstos deben ser para-

consistentes. Así, en Subrahmanian (1987), y Blair y Subrahmanian (en preparación, 1987a y 1987b) se introduce un sistema lógico paraconsistente, denominado lógica anotada, estableciendo una forma de programación paraconsistente. La lógica anotada empleada por estos autores fue construida sólo parcialmente, dado que su finalidad inicial era servir de base a una programación paraconsistente. La formulación de la lógica anotada como un sistema lógico completo es obra de da Costa, Subrahmanian y Vago. Ésta está siendo desarrollada por estos autores y por J. Abe. Se intenta desarrollar una lógica anotada de orden superior, una teoría de modelos anotada y una teoría de conjuntos anotada.

No es posible entrar aquí en los detalles técnicos del tema ya que son altamente matematizados. Lo único que debe quedar claro es que la programación anotada evidencia una enorme fuerza y simplicidad, encontrando las más variadas aplicaciones en numerosos sistemas expertos de la economía, la medicina, etc. Para una descripción de la programación paraconsistente, consultar da Costa y Subrahmanian (1989).

Técnicas paraconsistentes no anotadas importantes en programación, demostración automática de teoremas e inteligencia artificial son las de Carnielli (1987, 1990), Carnielli y Marques (1990) y Buschbaum y Pequeno (en prensa).

En síntesis, la manipulación de sistemas complejos de información, que generalmente aparecen por ejemplo en inteligencia artificial, sólo puede ser efectuada de modo natural y cómodo por medio de técnicas paraconsistentes. Quizás no sea una exageración afirmar que las máquinas del futuro serán, básicamente, paraconsistentes.

BIBLIOGRAFÍA

Alves, E. H. (1976), *Lógica e Inconsistência: Um Estudo dos Cálculos* C_n, $1 \leq n \leq \omega$, Universidade Estadual de Campinas, Brasil.

Arruda, A. I. (1964), *Considerações sobre os Sistemas Formais* NF_n, Universidade Federal do Paraná, Curitiba, Brasil.

Arruda, A. I. (1970a), «Sur les systèmes formels NF_i de da Costa»: *C. R. Acad. Sc. Paris, 270A,* 1.081-1.084.

Arruda, A. I. (1970b), «Sur le système NF_ω»: *C. R. Acad. Sc. Paris, 270A,* 1.137-1.139.

Arruda, A. I. (1974), «Le schéma de la séparation et les calculs J_n»: *Matematica Japonicae, 19,* 18-186.

Arruda, A. I. (1975), «Remarques sur les systèmes C_n»: *C. R. Acad. Sc. Paris, 280A,* 1.253-1.256.

Arruda, A. I. (1977), «On the imaginary logic of N. A. Vasilév», en Arruda, da Costa y Chuaqui (comps.), 1977.

Arruda, A. I. (1980), «A survey of paraconsistent logic», en Arruda, da Costa, Chuaqui (comps.), 1980.

Arruda, A. I. (1984), «N. A. Vasilév: A forerunner of paraconsistent logic»: *Philosophia Naturalis, 21,* 472-491.

Arruda, A. I. y Alves, E. H. (1979a), «Some remarks on the logic of vagueness»: *Bull. Sec. of Logic, Polish Acad. Sc., 8,* 133-138.

Arruda, A. I. y Alves, E. H. (1979b), «A semantical study of some systems of vagueness logic»: *Bull. Sec. of Logic, Polish Acad. of Sc., 8,* 139-144.
Arruda, A. I. y Alves, E. H. (1966), «O paradoxo de Curry-Moh Shaw-Kwei»: *Boletim da Sociedade Matemática de São Paulo, 18,* 83-89.
Arruda, A. I. y Alves, E. H. (1970), «Sur le schéma de la séparation»: *Nagoya Mathematical Journal, 38,* 71-84.
Arruda, A. I. y Alves, E. H. (1977), «Une semantique pour le calcul $C_1^{=}$»: *C. R. Acad. Sc. Paris, 284A,* 279-282.
Arruda, A. I. y Alves, E. H. (1984), «On the relevant systems P and P^* and some related systems»: *Studia Logica, 43,* 33-49.
Arruda, A. I. y da Costa, N. C. A. (1964), «Sur une hiérarchie de systèmes formels»: *C. R. Acad. Sc. Paris, 259,* 2.943-2.945.
Arruda, A. I., da Costa, N. C. A. y Chuaqui, R. (comps.) (1977), *Non-Classical Logic, Model Theory and Computability,* North-Holland, Amsterdam.
Arruda, A. I., da Costa, N. C. A. y Chuaqui, R. (comps.) (1980), *Mathematical Logic in Latin America,* North-Holland, Amsterdam.
Arruda, A. I., da Costa, N. C. A. y Sette, A. M. (comps.) (1980), *Proccedings of the Third Brazilian Conference on Mathematical Logic,* Sociedade Brasileira de Lógica, São Paulo.
Asenjo, F. G. (1966), «A calculus of antinomies»: *Notre Dame Journal of Formal Logic, 7,* 103-105.
Asenjo, F. G. y Tamburino, J. (1975), «Logic of antinomies»: *Notre Dame Journal of Formal Logic, 16,* 17-44.
Avron, A. (1990a), «Relevance and Consistency-a new approach»: *Journal of Symbolic Logic, 55,* 707-732.
Avron, A. (1990b), «Relevance and Paraconsistency-a new approach»: *Notre Dame Journal of Formal Logic, 31,* 169-202.
Batens, D. (1980), «A completeness proof method for extensions of the implicational fragment of the propositional calculus»: *Notre Dame Journal of Formal Logic, 21,* 509-517.
Batens, D. (1980), «Paraconsistent extensional propositional logics»: *Logique et Analyse, 90-91,* 195-234.
Bazhanov, V. A. (1989), *N. A. Vasilev,* Nauka, Moscu.
Belnap, N. (1977), «A useful four-valued logic», en Dunn y Epstein (comps.), 1977.
Blair, H. A. y Subrahmanian, V. S., *A logical framework for approximate reasoning in logic programming* (en preparación).
Blair, H. A. y Subrahmanian, V. S. (1987a), «Paraconsistent logic programming», en *Lecture Notes in Computer Science, 287,* 340-360.
Blair, H. A. y Subrahmanian, V. S. (1987b), «Paraconsistent generally Horn logic programming language: syntax and semantics», manuscrito.
Blok, W. y Pigozzi, D. (1989), «Algebraizable logics»: *Memoirs of the American Mathematical Society, 77,* 396.
Bunder, M. V. (1974), «Propositional and predicate calculs based on combinatory logic»: *Notre Dame Journal of Formal Logic, 15,* 25-34.
Bunder, M. V. (1980), «A new hierarchy of paraconsistent logic», en Arruda, da Costa y Sette (comps.), 1980.
Bunder, M. V. (1983), «On Arruda and da Costa's logics J_1 to J_5»: *The Journal of Non-Classical Logic, 2,* 43-48.
Bunder, M. V. (1989), «On paraconsistent combinatory logic», en Priest, Routley y Norman (comps.), 1989.
Buschbaum, A. y Pequeno, T., *Lógicas que son simultáneamente paraconsistentes y paracompletas* (en prensa).
Carnielli, W. A., *Inconsistent reasoning with consistent conclusions: a paraconsistent approach to machine thinking* (en prensa).

Carnielli, W. A. y Alcántara, L. P. (1984), «Paraconsistent algebras»: *Studia Logica, 43,* 79-88.

Carnielli, W. A. y Marques, M. L. (1990), «Reasoning under inconsistent knowledge», *Rapport Interne IRIT/90-15/R,* Université Paul Sabatier, Inst. de Recherche en Inf., Toulouse.

Costa, N. C. A. da (1958), «Nota sobre o conceito de contradição»: *Anuário da Soc. Paranaense de Matemática, 1,* NS, 6-8.

Costa, N. C. A. da (1959), «Observações sobre o conceito de existência em Matemática»: *Anuário da Sociedade Paranaense de Matemática, 2,* 16-19.

Costa, N. C. A. da (1963a), *Sistemas Formais Inconsistentes,* Universidade Federal do Paraná, Brasil.

Costa, N. C. A. da (1963b), «Calculs propositionels pour les systèmes formels inconsistents»: *C. R. Acad. Sc. Paris, 257,* 3.790-3.793.

Costa, N. C. A. da (1964a), «Calculs de prédicats pour les systèmes formels inconsistents»: *C. R. Acad. Sc. Paris, 258,* 27-29.

Costa, N. C. A. da (1964b), «Calculs de prédicats avec égalité pour les systèmes formels inconsistants»: *C. R. Acad. Sc. Paris, 258,* 1.111-1.113.

Costa, N. C. A. da (1964c), «Calculs de descriptions pour les systèmes formels inconsistants»: *C. R. Acad. Sc. Paris, 258,* 1.366-1.368.

Costa, N. C. A. da (1964d), «Sur un système inconsistant de théorie des ensembles»: *C. R. Acad. Sc. Paris, 258,* 3.144-3.147.

Costa, N. C. A. da (1965), «Sur les systèmes formels C_1, C_1^*, $C_1^=$, D_1 et NF_1»: *C. R. Acad. Sc. Paris, 260,* 5.427-5.430.

Costa, N. C. A. da (1966a), *Algebras de Curry,* São Paulo.

Costa, N. C. A. da (1966b), «Opérations non-monotones dans les treillis»: *C. R. Acad. Sc. Paris, 263A,* 429-432.

Costa, N. C. A. da (1967a), «Une nouvelle hiérarchie de théories inconsistants»: *Publications du Département de Mathématiques, Université de Lyon, 4,* 2-8.

Costa, N. C. A. da (1967b), «Filtres et idéaux d'une algèbre C_n»: *C. R. Acad. Sc. Paris, 264A,* 549-552.

Costa, N. C. A. da (1971), «Remarques sur le système NF_1»: *C. R. Acad. Sc. Paris, 272A,* 1.149-1.151.

Costa, N. C. A. da (1974a), «Remarques sur les calculs C_n, C_n^*, $C_n^=$ et D_n»: *C. R. Acad. Sc. Paris, 278A,* 819-821.

Costa, N. C. A. da (1974b), «On the theory of inconsistent systems»: *Notre Dame Journal of Formal Logic, 16,* 497-510.

Costa, N. C. A. da (1975), «Remarks on Jaśkowski's discussive logic»: *Reports on Mathematical Logic, 4,* 7-16.

Costa, N. C. A. da (1986), «On paraconsistent set theory»: *Logique et Analyse, 115,* 361-371.

Costa, N. C. A. da (1989), «Matemática e Paraconsistência», *Monografias da Soc. Paranaense de Matemática,* Curitiba, Brasil.

Costa, N. C. A. da y Alves, E. H. (1977), «A semantical analysis of the calculi C_n»: *Notre Dame Journal of Formal Logic, 18,* 621-630.

Costa, N. C. A. da y Carnielli, W. A. (1986), «On paraconsistent deontic logic», *Philosophia, 16,* 293-305.

Costa, N. C. A. da y Chuaqui, R., *The logic of quasi-truth* (en prensa).

Costa, N. C. A. da y Dubikajtis, L. (1968), «Sur la logique discoursive de Jaśkowski»: *Bull. Acad. Pol. Sc., 15,* 551-557.

Costa, N. C. A. da y Dubikajtis, L. (1977), «On Jaśkowski's discussive logic», en Arruda, da Costa y Chuaqui (comps.), 1977.

Costa, N. C. A. da y French, S. (1988), «Belief and contradiction»: *Critica, 20,* 3-11.
Costa, N. C. A. da y Marconi, D. (1986), «A note on paracomplete logic»: *Atti Accad. Lincei Rend. Ol. Sci. Fis. Mat., Natur, 8,* 80, n.os 7-12, 504-509.
Costa, N. C. A. da y Marconi, D. (1989), «An overview of paraconsistent logics in the 80s»: *The Journal of Non-Classical Logic, 1,* 5-31.
Costa, N. C. A. da y Sette, A. M. (1969), «Les algèbres C_ω»: *C. R. Acad. Sc. Paris, 268,* 1.011-1.014.
Costa, N. C. A. da y Wolf, R. G. (1980), «Studies in paraconsistent logic I: The dialectical principle of the unity of opposites»: *Philosophia, 9,* 189-217.
Costa, N. C. A. da y Wolf, R. G. (1985), «Studies in paraconsistent logic II: Quantification and the unity of opposites»: *Revista Colombiana de Matemáticas, 19,* 56-67.
Di Prisco, C. (comp.) (1985), *Methods in Mathematical Logic,* Springer, Berlin.
D'Ottaviano, I. M. L. (1982), *Sobre uma Teoria de Modelos Trivalentes,* Universidade de Campinas, Brasil.
D'Ottaviano, I. M. L. (1985a), «The completeness and compactness of a three valued first-order logic»: *Revista Colombiana de Matemáticas, 15,* 31-42.
D'Ottaviano, I. M. L. (1985b), «The model extension theorems for J_s-theories», en Di Prisco (comp.), 1985.
D'Ottaviano, I. M. L. (1990), «On the development of Paraconsistent Logic and da Costa's work»: *The Journal of Non-Classical Logic, 7,* 1/2, 9-72.
D'Ottaviano, I. M. L. y Costa, N. C. A. da (1970), «Sur un problème de Jaśkowski»: *C. R. Acad. Sc. Paris, 270,* 1.349-1.353.
Dunn, H. M. y Epstein, R. L. (comps.) (1977), *Modern Uses of Multiple-Valued Logics,* Reidel, Berlin.
Epstein, R. L. (1990), *The Semantics Foundations of Logic, Volume I: Propositional Logics,* Nijhoff International Philosophy Series, vol. 35, Kluwer.
Fidel, M. (1977), «The decidability of the calculi C_n»: *Reports on Mathematical Logic, 8,* 31-40.
Grana, N. (1983), *Lógica Paraconsistente,* Loffredo, Nàpoli.
Grana, N. (1990a), *Contradizione e Incompletezza,* Liguore, Nàpoli.
Grana, N. (1990b), *Sulla Teoria delle Valutazioni di N. C. A. da Costa,* Liguore, Nàpoli.
Heyting, A. (comp.) (1959), *Constructivity in Mathematics,* North-Holland, Amsterdam.
Karpenko, A. S. (1984), «Paraconsistent structure whithin many-valued logic», en *Many-valued, Relevant, and Paraconsistent Logics,* Logic Seminar, Institute of Philosophy, Acad. Sc. URSS, 39-39 (en ruso).
Kleene, S. C. (1952), *Introduction to Metamathematics,* Van Nostrand, Amsterdam.
Kotas, J. (1975), «Discussive sentential calculus of Jaśkowski»: *Studia Logica, 34,* 149-168.
Kotas, J. y Costa, N. C. A. de (1977), «On some modal systems defined in connection with Jaśkowski's problem», en Arruda, da Costa y Chuaqui (comps.), 1977.
Lewin, R. A., Mikenberg, I. F. y Schwarze, M. G. (1990), «Algebraization of Paraconsistent Logic P_1»: *The Journal of Non-Classical Logic, 7,* 145-154.
Lewin, R. A., Mikenberg, I. F. y Schwarze, M. G. (1991), «C_1 is not algebraizable»: *Notre Dame Journal of Formal Logic,* 32, 609-611.
Lewin, R. A., Mikenberg, I. F. y Schwarze, M. G. (1994), «P_1 algebras»: *Studia Logica, 53,* 21-28.
Lewin, R. A., Mikenberg, I. F. y Schwarze, M. G., «Algebraization of several deductive systems», manuscrito.
Lukasiewicz, J. (1910), «Über den Satz des Widerspruchs bei Aristoteles»: *Bull. Inter. de l'Académie des Sciences de Cracovie, Classe d'Histoire de Philosophie,* 15-38.
Lukasiewicz, J. (1971), «On the principle of contradiction in Aristotle»: *Review of Metaphysics, 24,* 485-509.

Marconi, D. (comp.) (1979), *La formalizzazione della Dialettica,* Rosenberg & Sellier, Torino.
Meinong, A. (1907), *Über die Stellung der Gegenstandstheorie im System der Wissenschaften,* R. Voitlander, Leipzig.
Mortensen, C. (1980), «Every quotient algebra for C_1 is trivial»: *Notre Dame Journal of Formal Logic, 21,* 694-700.
Mortensen, C. (1990), «Models for inconsistent and incomplete differential calculus»: *Notre Dame Journal of Formal Logic, 31,* 274-285.
Nelson, D. (1959), «Negation and separation of concept in constructive mathematics», en Heyting (comp.), 1959.
Peña, L. (1979), *Contradiction et Verité,* Université of Liège.
Peña, L. (1980), *Formalización y lógica dialéctica,* Universidad Católica de Ecuador.
Pinter, C. (1980), «The logic of inherent ambiguity», en Arruda, da Costa y Sette (comps.), 1980.
Priest, G. (1979), «The logic of paradox»: *Journal of Philosophical Logic, 8,* 219-241.
Priest, G., Routley, R. y Norman, J. (comps.) (1989), *Paraconsistent logic. Essays on the inconsistent,* Philosophia Verlag, München.
Puga, L. P. (1985), *Uma lógica do Querer,* Universidade Católica de São Paulo.
Puga, L. P., Costa, N. C. A. da y Carnielli, W. A. (1993), «Kantian and non-kantian logic»: *Logique et Analyse, 121-122,* 3-9.
Raggio, A. R. (1968), «A propositional sequence-calculi for inconsistent systems»: *Notre Dame Journal of Formal Logic, 9,* 359-366.
Rescher, N. y Brandon, R. (1980), *The Logic of Inconsistency,* Basil Blackwell, Oxford.
Routley, R. y Meyer, R. K. (1976), «Dialectical logic, classical logic and the consistency of the world»: *Studies in Soviet Thought, 16,* 1-25.
Routley, R., Meyer, R. K., Plumwood, V. y Brady, R. (1983), *Relevant Logics and their Rivals,* Ridgeview.
Sette, A. M. (1973), «On the Propositional Calculus P^1»: *Mathematicae Japonicae, 16,* 173-180.
Subrahmanian, V. S. (1987), «On the semantics of quantitative logic programs», en *Proceedings of the 4th IEEE Symposium on Logic Programming,* Computer Society Press, Washington DC, 173-182.
Urbas, I. (1989), «Paraconsistency and the *C*-systems of da Costa»: *Notre Dame Journal Formal Logic, 30,* 583-597.
Vasilev, N. A. (1925), «Imaginary (non-aristotelica) logic», *Atti dei V Congresso Internazionale di Filosofia,* Nàpoli, 107-109.

LOGICA EPISTÉMICA

Max A. Freund

INTRODUCCIÓN

Nuestro discurso sobre el conocimiento y creencia adquiere diversas perspectivas y niveles de generalidad. Podemos establecer, por ejemplo, relaciones entre los mecanismos biológicos del ser humano y el conocimiento de éste, concentrarnos en la fundamentación del conocimiento matemático o ligar los elementos psicológicos a las creencias cosmológicas. Este tipo de discurso ha sido denominado «epistémico» y es la forma de discurso presente en este artículo. Sin embargo, no todas las perspectivas y niveles de generalidad de este discurso se desarrollarán aquí. No trataremos, por ejemplo, las dimensiones filosóficas o sociopsicológicas ya apuntadas.

La dimensión de nuestro discurso interpretará la entidad presupuesta en los conceptos de conocimiento y creencia, esto es, aquello que conoce o cree (el cual llamaremos «el agente»), de una manera muy general. Este enfoque, entonces, no se concentrará solamente en un agente con ciertas características definidas (como podría ser un ser humano), sino que contemplará las diferentes posibilidades que ofrece el uso de los conceptos de creencia y conocimiento. De este modo, nuestra dimensión permitirá identificar, en ciertos casos, aquello que conoce o cree como un ser humano particular y, en otros, como un grupo social, individuo ideal, computador, serie de computadores o, en forma más general, como un autómata o serie de autómatas[1].

Nuestro discurso toma como punto de partida la posibilidad misma de razonar relativo a un agente o agentes dados. Este punto de partida considera el hecho patente de que ciertos razonamientos, dentro de contextos epistémicos, no *parecen* regirse por las lógicas clásicas proposi-

1. Cf. Hopcroft y Ullman (1979) para detalles sobre teoría de autómatas.

cional o de primer orden con identidad[2]. Recuérdense dos principios básicos en estas lógicas: la regla de reemplazo de oraciones equivalentes y la ley de Leibniz de substitución de los idénticos[3].

Debería ser posible, de acuerdo con la ley de Leibniz, argumentar que (C) José Rubí sabe que César Augusto Sandino fue mandado asesinar por el padre del dictador nicaragüense que murió en Paraguay, a partir de las afirmaciones: i) José Rubí sabe que César Augusto Sandino fue mandado asesinar por el padre de Anastasio Somoza Debayle, y ii) Anastasio Somoza Debayle es el dictador nicaragüense asesinado en Paraguay. Por otra parte, la regla de reemplazo debería justificarnos en la siguiente inferencia: i*)
siva. ii*)
máquina de registros ilimitados. Por lo tanto, (C*) Julio Ramírez cree que una función computable es computable por una máquina de registros ilimitados.

Los razonamientos antes descritos, sin embargo, no pueden ser considerados válidos, pues es posible que (i), (ii), (i*)
sen verdaderas, pero (C) y (C*)
ejemplo, que Rubí no supiera a dónde se fue Somoza al huir de Nicaragua o si murió. En el segundo caso, podría ser que los conocimientos de Julio sobre funciones recursivas se hubiesen limitado a un curso introductorio de teoría de computabilidad en donde el tema de los registros ilimitados nunca fue considerado.

Hay argumentos, por lo tanto, que no parecen obedecer a principios importantes de la lógica clásica. Una manera de explicar estas anomalías ha sido suponer que los operadores de la forma «δ sabe que β» y «δ cree que β» son, en realidad, relaciones sintácticas de primer orden, esto es, en forma más precisa, relaciones entre individuos y nombres de oraciones. Si S es una oración, designaremos el nombre de esta oración con «S». De acuerdo con esa propuesta, entonces, las oraciones (i*)
(i) deberían ser expresadas, respectivamente, como Cree (Julio, «una función computable es recursiva») y Sabe (José Rubí, «César Augusto Sandino fue mandado asesinar por el padre de Anastasio Somoza Debayle»). Esta interpretación mostraría, evidentemente, por qué los principios lógicos de primer orden mencionados son inaplicables: en el primer razonamiento, porque el término «Anastasio Somoza Debayle» en (ii) y en (i) constituirían diferentes nombres; en el segundo, porque la oración «una función computable es recursiva» en (i*)
un nombre.

La propuesta sintáctica permitiría tratar principios relativos a operadores, ligados a los conceptos de creencia y conocimiento, como teorías de primer orden. Sin embargo, esta propuesta, sugerida, por ejemplo,

2. Cf. Quesada, 1994.
3. *Ibid.*

en Quine (1975), ha presentado varios problemas. Entre éstos, el fundamental, el cual se sigue de resultados expresados en Montague (1963) y Thomason (1980), es que cualquier teoría de primer orden que contenga niveles de aritmética suficientes para permitir expresar la sintaxis de la teoría y ciertos principios epistémicos relativos a las relaciones Sabe (δ, «ß») o Cree (δ, «ß») es inconsistente[4]. Estos principios, sin embargo, son de amplia aceptación.

Por los problemas apuntados, nuestro enfoque no se orientará dentro de la interpretación sintáctica. Asumirá, más bien, que los operadores de la forma «δ sabe que ß», y «δ cree que ß», así como muchos otros ligados a los conceptos de conocimientos y creencia, hacen inoperantes algunos de los principios de la lógica clásica[5]. El estudio de las propiedades lógicas de esos operadores o, en forma más precisa, el análisis lógico-formal del razonamiento sobre conocimiento y creencia se denomina «lógica epistémica»[6]. La dimensión del discurso epistémico en el presente artículo se centrará en este análisis. Sin embargo, no intentaremos hacer una compilación de la literatura existente en lógica epistémica. Preferiremos referir al lector a ciertas obras que cumplen esta función, tales como Lenzen (1978), y Baeuerle y Cresswell (1989). Aquí nos ocuparemos, más bien, de describir, en forma general, los métodos que ha tomado tal lógica en su objetivo de establecer principios y condiciones bajo las cuales se da la relación de consecuencia lógica en contextos epistémicos.

I. LENGUAJES FORMALES EPISTÉMICOS

Como análisis formal que es, la lógica epistémica enuncia los principios generales de consecuencia lógica, en contextos epistémicos, en relación con lenguajes formales. Diversos niveles de complejidad con respecto a estos lenguajes han sido explorados. Se han considerado lenguajes muy simples como los proposicionales así como más complejos como los de primer orden o de órdenes superiores, a los cuales se han agregado operadores epistémicos[7].

En el caso de lenguajes proposicionales, encontramos que contienen un conjunto enumerable de proposiciones atómicas, P_1, P_2... y un conjunto de constantes lógicas $\rightarrow$, $\neg$, (,). Estas constantes han de interpretarse, intuitiva y respectivamente, como la implicación, negación, paréntesis izquierdo y paréntesis derecho. El lenguaje ha de contener también un conjunto finito de operadores proposicionales $O_1 \ldots O_n$, cuya

4. Sobre cómo la aritmética puede expresar la sintaxis de una teoría, cf. Mosterín, 1994.
5. Sin embargo, para desarrollos ulteriores de la propuesta sintáctica, cf. Hass, 1986; Asher y Kamp, 1986; Koons, 1988; y Perlis, 1988.
6. La idea de una lógica epistémica se remonta a von Wright.
7. Para detalles sobre lenguajes y lógicas de órdenes superiores, cf. Jané, 1994.

interpretación intuitiva los constituye en operadores epistémicos. Lo que es una fórmula bien formada (fbf) es definida en estos lenguajes de la siguiente manera: I) toda variable proposicional es una fbf; II) si σ y α son fbfs entonces $\neg\sigma$, $(\sigma\rightarrow\alpha)$, $O_m\sigma$ $(m\leq n)$ también lo son. Como ejemplo, podemos citar el lenguaje que tiene como operadores $S_0...S_n$ (para algún número natural n) y en donde «$S_i\sigma$» $(i\leq n)$ es interpretado intuitivamente como «el agente i sabe que σ». Otros operadores epistémicos que han sido expresados en lenguajes proposicionales incluyen «α cree que ß», «es cognoscible que ß», «es cognoscible en forma constructiva que ß», «ß puede ser probado» y «ß es de conocimiento común». Nótese que no siempre un operador tiene que hacer explícito el agente, como sucede en los dos últimos operadores.

En el caso de lenguajes epistémicos de primer orden, encontramos que contienen, además de las constantes lógicas de lenguajes proposicionales y operadores epistémicos, un conjunto enumerable de variables de individuos x_1, x_2..., un conjunto enumerable de constantes de individuos a_1, a_2, a_3..., así como un conjunto enumerable de predicados y otro de funciones de cualquier número de argumentos P_1, P_2, P_3,... y f_1, f_2...; y un operador $\forall$ (llamado el cuantificador universal). Podemos extendernos a lenguajes de órdenes superiores agregando, como es bien conocido, un conjunto enumerable de variables por cada orden.

Cómo hemos de definir los que es una fbf dentro de lenguajes de primer o de órdenes superiores se vuelve problemático, fundamentalmente por las objeciones expuestas en Quine (1971) respecto a la cuantificación dentro de contextos intensionales. Por razones de espacio no profundizaremos en estas objeciones y las respuestas a éstas. Preferimos referir al lector a Lenzen (1978) para un resumen de la situación. Sí podemos hacer notar que nuestra actitud ante estas objeciones determinará si hemos de asumir, por ejemplo, en el caso de lenguajes de primer orden epistémicos, la cláusula (IIIa): si σ es una fbf, entonces $(\forall x)\sigma$ es una fbf; la cláusula (IIIb): si σ es una fbf no epistémica (esto es, sin operadores epistémicos), entonces $(\forall x)\sigma$ es una fbf; o la cláusula (IIIc): si σ es una fbf que no es a la vez abierta (esto es, con variables libres) y epistémica, entonces $(\forall x)\sigma$ es una fbf. En otros términos, de acuerdo con nuestras sensibilidades, podríamos aceptar o no que se cuantifique dentro de contextos epistémicos como podría ser $(\forall x)\neg S_nF_x$ (para algún $n\in N$), en el caso de un lenguaje de primer orden con los operadores $S_1...S_n$ antes mencionados.

Designemos con LE-$\{O_1...O_n\}$ un lenguaje (ya sea proposicional, de primer o de orden superior) con operadores $O_1...O_n$. Como hicimos notar anteriormente, se parte de una interpretación intuitiva epistémica de estos operadores, así como en LE-$\{S_1...S_n\}$, por ejemplo, partimos de la interpretación intuitiva de $S_i\sigma$ $(i\leq n)$ como el «agente i sabe que σ». La lógica epistémica puede proceder en dos direcciones (no excluyentes) de análisis lógico de operadores epistémicos: una, que llamaremos «sintáctica», y la otra «semántica-formal».

En la dirección sintáctica se contempla un subconjunto Ax de fbfs del lenguaje (los axiomas) y un conjunto R de reglas de transformación (las reglas de inferencia). Obtenemos de este modo un sistema formal axiomático, esto es, un triple $<L\text{-}\{O_n \ldots O_n\}, Ax, R>$. La idea general es que el sistema formal capture los aspectos lógicos esenciales de los operadores.

A manera de ejemplo, consideremos el lenguaje proposicional LE-$\{S_1 \ldots S_n\}$, para algún número natural n. Relativo a este lenguaje, podemos formular un sistema formal, el cual llamaremos «el sistema H». Sean σ y α fbfs del lenguaje e $i \leq n$. H está constituido por las siguientes reglas y esquemas de axiomas[8].

Axiomas

A1. Todas las tautologías proposicionales,
A2. $(S_i\sigma \ \& \ S_i(\sigma \rightarrow \delta)) \rightarrow S_i\delta$,
A3. $S_i\sigma \rightarrow \sigma$,
A4. $S_i\sigma \rightarrow S_iS_i\sigma$,
A5. $\neg S_i\sigma \rightarrow S_i \neg S_i\sigma$;

Reglas de inferencia

R1. si σ, $\sigma \rightarrow \alpha$, entonces α
R2. si α, entonces $S_i\alpha$.

A1 y R1 son elementos de la lógica proposicional clásica. A2 afirma que el conocimiento de un agente está cerrado bajo la implicación y R2 que está cerrado bajo las deducciones del sistema. A3 expresa la idea clásica de que se conocen sólo verdades. A4 y A5 son axiomas de introspección: el agente puede contemplar su conocimiento y sabrá lo que él conoce y no conoce. Es importante hacer notar que se ha discutido ampliamente, dentro de círculos filosóficos, sobre lo apropiado de este sistema como una formalización del concepto «α sabe que ß». El lector puede consultar Lenzen (1978) para una guía completa sobre estas discusiones. Sin embargo, queremos mencionar que A2 y R2 son los elementos que más controversia han causado, pues nos fuerzan a concebir el agente como un cognoscente ideal: un agente que conoce todas las fórmulas válidas así como todas las consecuencias lógicas de su conocimiento. Esto, obviamente, no va de acuerdo con una interpretación del agente como un ser humano o como un computador limitado por tiempo y espacio en la memoria que puede usar. Sin embargo, como mencionaremos más adelante, existen otros tipos de agentes para los cuales el sistema H sí es apropiado.

8. H constituye uno de los sistemas formales más importantes de lógica epistémica. Fue enunciado en Hintikka (1962) y ha servido de referencia a investigaciones y desarrollos posteriores.

Aparte de la dirección sintáctica, podemos proceder a construir una semántica que interpreta el lenguaje formal epistémico. Se persigue con esto que la semántica constituya un modelo, en términos de conjuntos, de nuestras intuiciones relativas a los operadores epistémicos. Con esto obtenemos, a la vez, una caracterización de los aspectos lógicos de los operadores epistémicos. Dado esto, se explora la posibilidad de enunciar un sistema formal F tal que toda fbf σ caracterizada como una verdad lógica en la semántica sea un teorema en F y viceversa. Sin embargo, no podemos esperar que esto último sea siempre posible. Podría ser el caso que semánticas de segundo orden epistémicas siguieran la suerte de las estándar de segundo orden.

II. SEMÁNTICA DE MUNDOS POSIBLES

Las primeras semánticas formales de lógica epistémica fueron desarrolladas utilizando la noción de mundo posible, esto es, la idea intuitiva de que, además del presente mundo, existen otras formas o maneras en que el mundo pudo haber sido[9]. De estas semánticas iniciales, sólo expondremos, por razones de espacio, la primera formulada para lenguajes epistémicos proposicionales o de primer orden con operadores de la forma «α sabe que ß»[10]. La importancia de ésta radica en que ha servido de referencia para los desarrollos semánticos posteriores. El lector puede consultar, por ejemplo, Halpern y Moses (1984) así como Baeuerle y Cresswell (1989) para semánticas de otros operadores.

La semántica parte de la idea de que la información que un agente posee no le permite decidir cuál de los mundos (que él considera posibles) describe el mundo como es actualmente. Sobre la base de esta imagen, se interpreta intuitivamente el que un agente sepa que σ como «σ es verdadera en todos los mundos que el agente considera posibles (dada su información actual)». Por ejemplo, dada cierta información, el agente puede considerar posibles dos tipos de mundos, unos en donde no hay tuberculosis ni cáncer y otros donde no hay tuberculosis, pero sí hay cáncer. Siguiendo la interpretación intuitiva, se diría entonces que el agente sabe que la tuberculosis es capaz de ser eliminada, pero su información no le permite saber si el cáncer puede ser prevenido.

9. Para diversas concepciones sobre la naturaleza de los mundos posibles, cf. Loux, 1979. Alternativo a la noción de mundo posible, también se utiliza la noción de situación o escenario. Para esta última, cf. Barwise y Perry, 1983.

10. La primera exposición de este tipo de semántica se encuentra en Hintikka, 1962. Una formulación alternativa, históricametne importante, constituye la de R. Montague y que difiere de la expuesta en este apartado, entre otros aspectos, por la exclusión de la relación de accesibilidad. También, esta semántica, a diferencia de la de Hintikka, se ubica dentro de un proyecto lógico-lingüístico más amplio, cuyo propósito, entre otros, es el desarrollo de una lógica intencional general. Para detalles, cf. Anderson, 1984 y Partee, 1976.

Las ideas antes expresadas han sido formalizadas en términos de estructuras de Kripke[11]. Entenderemos por una estructura E de Kripke un tuplo $<M, \pi, P_1,...P_m>$, en donde M es un conjunto de mundos posibles, π es una asignación de valores de verdad a las proposiciones atómicas para cada mundo posible $m \in M$, de tal modo que $\pi(m, p) \in$ {Verdad, Falsedad} para cada $m \in M$ y proposición atómica p, P_i es una relación sobre M (para i = 1,..., m). P_i es la relación de posibilidad de i (llamada, también, la relación de accesibilidad de i). De este modo, hemos de entender intuitivamente que $(s, m) \in P_i$, esto es, m es accesible de acuerdo con la relación P_i desde s, como el agente i considera a m posible dada la información que posee en s.

Caracterizamos ahora la verdad de una fbf en una estructura y mundo m (en símbolos, $E, m \vdash \sigma$) como sigue:

$E, m \vdash p$ sii $\pi(m, p) = V$, para toda proposición atómica.
$E, m \vdash \neg\sigma$ sii no es el caso que $E, m \vdash \sigma$.
$E, m \vdash \sigma \rightarrow \delta$ sii $E, m \vdash \neg\sigma$ o $E, m \vdash \delta$
$E, m \vdash S_i\sigma$ sii $E, r \vdash \sigma$ para todo r tal que $(m, r) \in Pi$.

Esta semántica puede ser extendida para lenguajes de primer orden epistémicos. En este caso agregamos un dominio de objetos D y una función F que asigna diversas extensiones a los predicados en cada mundo posible, esto es, para todo $m \in M$ y predicado P de n argumentos, F(P, m) es un subconjunto de D^n.

De acuerdo con las propiedades de las relaciones de accesibilidad P_i, se obtienen nociones de verdad lógica que formalizan diversas interpretaciones intuitivas del operador «i sabe que ß». Por ejemplo, si se asume que P_i es reflexivo, simétrico y transitivo, entonces la noción de verdad lógica que proporciona esta semántica puede ser caracterizada por el sistema formal H. Esto es, los axiomas de H se constituirían en esquemas de verdades lógicas y las reglas nos llevarían de verdades lógicas a verdades lógicas. Por otra parte, toda verdad lógica en la semántica y regla que preservara verdad lógica podría ser demostrada en el sistema H. Si estipuláramos, por otra parte que P_i no fuese necesariamente simétrico, entonces perderíamos a A5 como un esquema de verdad lógica en la semántica resultante.

Independientemente de las propiedades que las relaciones de accesibilidad pueden poseer, cualquier semántica de este tipo siempre justificará A2 y R2. Por lo tanto, esta semántica nos obliga a concebir al agente como una entidad capaz de conocer todas las consecuencias lógicas de su conocimiento, lo cual la hace inadecuada como un modelo de agentes tales como seres humanos o computadores con limitaciones de espacio y tiempo en su memoria[12]. Este problema, denominado el problema de

11. Para detalles sobre estas estructuras, cf. Orayen, 1994.

12. Es importante notar, sin embargo, que la semántica es apropiada para representar sistemas distribuidos, esto es, colecciones de procesadores conectados por una red de comunicación. En este caso, los agentes serían los procesadores. Para detalles, cf. Halpern y Moses, 1985.

la omnisciencia lógica, ha estimulado el desarrollo de semánticas alternativas.

Dentro de la semántica misma de mundos posibles, se ha sugerido, por ejemplo, en Cresswell (1973) y Rantala (1982) introducir mundos no estándar, esto es, mundos que no cumplen, en forma general, con las leyes de la lógica clásica. Ejemplos de este tipo de mundos podrían ser aquellos en donde ciertas proposiciones no son ni verdaderas ni falsas o en donde determinadas proposiciones son verdaderas y falsas. El problema con este enfoque radica en que ese tipo de mundos no ha sido motivados en forma satisfactoria. Sin embargo, cf. Hintikka (1975) para una posible motivación. Es importante mencionar que mundos no estándar han permitido, en Levesque (1984), desarrollar una semántica que logra distinguir entre conocimiento implícito y explícito, lo cual ofrece un escape al problema apuntado: se posee sólo conocimiento implícito de todas las consecuencias lógicas de nuestro conocimiento, pero no así explícito.

Se han propuesto también enfoques alternativos al de los mundos posibles. El más prometedor de éstos ha sido el expuesto en Konolige (1985). La idea intuitiva detrás de esta propuesta consiste en lo siguiente: sea A un agente determinado, al cual asociamos un conjunto de oraciones $ß_A$, las cuales llamaremos las creencias básicas de A, y un conjunto de reglas de inferencia R_A. Ahora diremos que A sabe que σ si y sólo si A puede inferir σ a partir de $ß_A$ y R_A. Esta idea permitiría solucionar el problema de la omnisciencia lógica: el conjunto de reglas R_A no ha de coincidir necesariamente con un conjunto completo de reglas de deducción y esto posibilitaría el que A no conociera ciertas consecuencias lógicas de su conocimiento.

IV. CONSIDERACIONES FINALES

Existen varios temas de interés filosófico relacionados con la lógica epistémica que no hemos tratado hasta el momento. Por su desarrollo e importancia relativa, nos interesa mencionar ahora el de la matemática[13] epistémica.

Partamos de la idea de que el agente posee ciertas capacidades o niveles de información que le permiten adquirir un grado importante de conocimiento matemático. Por otra parte, asumamos que el agente no se encuentra limitado por aspectos de factibilidad tales como tiempo, memoria y desarrollo tecnológico. Podemos preguntarnos ahora cuáles limitaciones cognoscitivas tendría, en general, tal agente. La solución a este problema ha estimulado la construcción de diversos sistemas formales matemáticos, en los cuales la lógica subyacente es epistémica. De acuerdo con la

13. Por razones de espacio no expondremos los otros temas. El lector encontrará varios de éstos en las obras citadas.

lógica o la base matemática asumida, los sistemas varían. Esto es, relativo al tipo de agente que contemplemos, podemos admitir o rechazar ciertas formas de razonamiento y teorías matemáticas. Diversos agentes contemplados han justificado la introducción de operadores tales como «puede ser probado en principio que p», «puede ser decidido en forma algorítmica que p» y «es cognoscible en forma constructiva que p». Los diferentes tipos de agentes asumidos han justificado la introducción de teorías matemáticas como la aritmética de Peano, la teoría de conjuntos finitos o diversas teorías lógicas de conjuntos.

Lo interesante de los sistemas de matemática epistémica es que han permitido establecer algunos teoremas limitativos relativos a ciertas formas de conocimiento. Esto ha de tener una influencia futura en teorías epistemológicas sobre la matemática y disciplinas afines. El lector puede consultar Shapiro (1985), Freund (1991) y Reinhardt (1986) en donde se exponen así cómo se justifican varios de esos sistemas y teoremas limitativos.

Hemos descrito en forma muy general los aspectos importantes del enfoque lógico-formal del discurso epistémico. Hemos visto que la lógica epistémica primero construye lenguajes formales con operadores, cuya interpretación intuitiva los constituye en operadores epistémicos. Luego, busca la formulación de sistemas y semánticas formales que capturen los aspectos lógicos de esa interpretación. Gran parte de los trabajos en lógica epistémica se han concentrado en la formulación de sistemas formales. Falta desarrollar más la dirección de semánticas formales. Dentro de esta dirección creemos que surgirán los problemas y soluciones más interesantes.

BIBLIOGRAFÍA

Anderson, A. (1984), «General Intentional Logic» en Gabbay y Guenthner, 1984.

Asher, N. y Kamp. H. (1986), «The Knowers Paradox and Representational Theories of Attitudes», en Halpern, 1986, 131-148.

Baeuerle, R. y Cresswell, M. (1984), *Situations and Attitudes,* MIT Press, Cambridge.

Cresswell, M. (1973), *Logics and Languages,* Methuen and Co., London.

Freund, M. (1991), «Consideraciones lógico-epistémicas sobre una forma de conceptualismo ramificado»: *Crítica, 69.*

Gabbay, D. y Guenthner, F. (comps.) (1984), *Handbook of Philosophical Logic. Extensions of Classical Logic,* D. Reidel, Dordrecht.

Gabbay, D. y Guenthner, F. (comps.) (1989), *Handbook of Philosophical Logic. Topics in the Philosophy of Language,* D. Reidel, Dordrecht.

Halpern, J. (1986), *Theoretical Aspects of Reasoning about Knowledge,* Morgan Kaufman, Los Altos, CA.

Halpern, J. y Moses, Y. (1984), «A guide to the Modal Logics of Knowlege and Belief: A preliminary report», *Proceedings of the 9th International Joint Conference on Artificial Intelligence,* 480-490.

Hintikka, J. (1962), *Knowledge and Belief,* Cornell University P., Ithaca, New York.

Hintikka, J. (1975), «Impossible Possible Worlds Vindicated»: *Journal of Philosophical Logic, 4,* 475-484.

Hobbs, J. y Moore, R. (comps.) (1985), *Formal Theories of the Commonsense World,* Norwood, N. J.
Hopcroft, J. y Ullman, J. (1979), *Introduction to Automata Theory, Languages, and Computation,* Addison-Wesley.
Jané, I. (1994), «Lógica de orden superior», en el presente volumen, 105-128.
Konolige, K. (1985), «Belief and Incompleteness», en Hobbs y Moore (comps.), 1985, 259-404.
Koons, R. (1988), «Doxastic Paradoxes Without Self-Reference», en Vardi, 1988, 29-42.
Lenzen, W. (1978), «Recent Work of Epistemic Logic»: *Acta Philosophica Fennica, 30.*
Levesque, J. (1984), «A Logic of Implicit and Explicit Belief, *Proceedings of the National Conference on Artificial Intelligence,* 155-212.
Loux, J. (comp.) (1979), *The Possible and the Actual, Readings in the Metaphysics of Modality,* Cornell University Press, Ithaca, New York.
Montague, R. (1963), «Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability»: *Acta Philosophica Fennica, 16,* 153-167.
Mosterín, J. (1994), «Computabilidad», en el presente volumen, 271-288.
Orayen, R. (1994), «Lógica modal», en el presente volúmen, 289-322.
Partee, B. (comp.) (1976), *Montague Grammars,* Academic Press, New York.
Perlis, D. (1988), «Language with Self-Reference II: Knowledge, Belief, and Modality»: *Artificial Intelligence, 34,* 179-212.
Quesada, D. (1994), «Lógica clásica de primer orden», en el presente volumen, 71-104.
Quine, W. (1975), *The Ways of Paradox,* Harvard University Press, Cambridge, Mass.
Rantala, V. (1982), «Impossible World Semantics and Logical Omniscience»: *Acta Philosophica Fennica, 35,* 106-115.
Reinhardt, W. (1986), «Epistemic Theories»: *Journal of Philosophical Logic, 15,* 427-474.

LÓGICA TEMPORAL

Margarita Vázquez Campos

I. INTRODUCCIÓN

Los antecedentes de lo que hoy conocemos como «lógica temporal» o «lógica del tiempo» son casi tan antiguos como la misma lógica. Se suele citar como primer antecedente a Aristóteles, quien en el «Perí Hermeneias» escribe que el tiempo puede modificar el valor de verdad de las proposiciones. Aristóteles no desarrolla esta idea. Es Diodoro de Cronos, perteneciente a la Escuela Megárica, quien da los primeros pasos en esta disciplina al tratar el tema de la definición de las modalidades. Las ideas de Diodoro tuvieron una gran difusión en toda la lógica antigua y medieval.

Ya en nuestro siglo, la figura fundamental dentro de este campo es Arthur Prior. De hecho, Prior es el padre de la lógica temporal como una rama autónoma dentro de la lógica. Su idea básica es que el tiempo (el momento del tiempo) al que se refiere una proposición es decisivo para atribuir un valor de verdad a dicha proposición. En sus múltiples libros y escritos, Prior, que parte de una lógica temporal muy fuertemente vinculada a la modal, defiende la importancia de una lógica temporal autónoma (frente a los opositores a esta idea, como Quine) y ofrece multitud de axiomatizaciones de sistemas de lógica temporal, dependientes de las concepciones que se tengan en cada caso acerca del tiempo. Es decir, un sistema que recoja una concepción del tiempo circular será diferente de un sistema que, por ejemplo, recoja una concepción del tiempo como ramificado en el futuro.

A partir de los años 50, que es de donde datan las primeras aportaciones de Prior, el desarrollo de la lógica temporal ha sido enorme, especialmente en las dos últimas décadas. Esto es no sólo debido al interés formal intrínseco de este tipo de sistemas, sino también a la amplia variedad de campos a los que se puede aplicar. Burgess (1984) señala cinco

tipos diferentes de motivos que justifican el desarrollo de la lógica temporal: filosóficos, exegéticos, lingüísticos, informáticos y matemáticos. Desde el momento en que Burgess planteaba esto ha habido muchos desarrollos, sobre todo en lingüística y en informática.

En lingüística nos encontramos con la utilización de la lógica temporal para el estudio de las estructuras lógicas de las formas de razonamiento asociadas con el lenguaje natural (Bras, 1990). Estos modos de razonamiento pueden llevar, además, al desarrollo de importantes aplicaciones prácticas. En este campo, es especialmente destacable el cálculo de eventos y el razonamiento por defecto (especialmente interesante es Kowalski y Sergot, 1986).

En informática, la lógica temporal se ha manifestado como especialmente apta para ser aplicada a la teoría de la programación, especialmente para el estudio del comportamiento tanto de programas secuenciales como de programas paralelos (Pnueli, 1977; Audureau, Enjalbert y Fariñas del Cerro, 1989; Bahsoun, 1988). En este campo, los sistemas pueden ser de lógica temporal lineal o ramificada. Aquí, tras el estudio de los formalismos, se pueden definir métodos de deducción automática para ellos y realizar máquinas de inferencia abstracta que los soporten. En esta línea, cabe señalar que la semántica de mundos posibles, utilizada en lógica modal y heredada con ciertas modificaciones en lógica temporal, puede ser sustituida por una semántica basada en autómatas (por ejemplo, Thayse, 1989).

El tipo de lógica temporal que aquí se va a tratar se corresponde con lo que se ha llamado lógica del tiempo gramatical. Hay otro tipo de lógica del tiempo llamada lógica cronológica, que en lugar de introducir nuevos operadores para el pasado y el futuro, está basada en relaciones temporales. Para una lógica de este tipo, véase Peña (1989).

II. AXIOMATIZACIÓN DE LA LÓGICA TEMPORAL

Como se ha dicho en el apartado anterior, la axiomatización de los sistemas de lógica temporal va a depender de la concepción que se tenga del tiempo. Una axiomatización constará de un conjunto de axiomas de la lógica proposicional clásica y de los axiomas necesarios para reflejar las propiedades del tipo de tiempo. En algunos casos, será necesario introducir axiomas de la lógica modal (por ejemplo, en un sistema ramificado en el futuro, pero cuya ramificación provenga de la modalidad y no de la temporalidad). Las principales propiedades que el tiempo puede tener, quedarían recogidas en los siguientes axiomas:

Ax. 1. $G(A \rightarrow B) \rightarrow (GA \rightarrow GB)$
Ax. 2. $H(A \rightarrow B) \rightarrow (HA \rightarrow HB)$
Ax. 3. $A \rightarrow HFA$
Ax. 4. $A \rightarrow GPA$

Ax. 5. GA, cuando A es una tautología de la lógica proposicional clásica
Ax. 6. HA, cuando A es una tautología de la lógica proposicional clásica
Ax. 7. $FFA \to FA$
Ax. 8. $(FA \& FB) \to (F(A \& B) \vee (F(A \& FB) \vee F(FA \& B)))$
Ax. 9. $(PA \& PB) \to (P(A \& B) \vee (P(A \& PB) \vee P(PA \& B)))$
Ax. 10. $GA \to FA$
Ax. 11. $HA \to PA$
Ax. 12. $FA \to FFA$
Ax. 13. $GA \to A$
Ax. 14. $GA \to HA$

En estos axiomas,
& y v, encontramos nuevos operadores monarios: G, H, F y P. G y F son los operadores relativos al futuro y H y P los del pasado. F puede ser entendido como «será el caso que»,
A partir de ellos,
$H =_{df} \neg P \neg$. La similitud con los operadores M y L de lógica modal es evidente.

Entre las propiedades del tiempo que reflejarían estos axiomas,
mos ver que el Ax. 7 refleja la transitividad,
la derecha (futuro),
el Ax. 10 la infinitud en el futuro,
el Ax. 12 la densidad,
Estas propiedades habrán de quedar igualmente expresadas en la semántica.

III. SEMÁNTICA DE LA LÓGICA TEMPORAL

La semántica de la lógica temporal está basada en la noción de momento histórico (W es el conjunto de momentos). Entre estos momentos se da una relación de ulterioridad R (o relación antes/después).

Así,
del tipo $<W,R,v>$ donde,

i) $W \neq \emptyset$

ii) $R \subseteq W^2$ y es una relación cuyas propiedades dependerán de la concepción del tiempo.

iii) Siendo F el conjunto de todas las fórmulas bien formadas (f.b.f.), $v: FxW \to \{1,$
$w_i, w_j \in W$, $A,B \in F$ y variable proposicional p:

I) $v(p,w_i) = 1$ ó $v(p, w_i) = 0$
II) $v(A \to B,w_i) = 1$ si y sólo si $v(A,w_i) = 0$ ó $v(B,w_i) = 1$
III) $v(\neg A,w_i) = 1$ si y sólo si $v(A,w_i) = 0$

IV) $v(PA,w_i)=1$ si y sólo si para algún w_j, tal que w_jRw_i, $v(A,w_j)=1$

V) $v(FA,w_i)=1$ si y sólo si para algún w_j, tal que w_iRw_j, $v(A,w_j)=1$

VI) $v(HA,w_i)=1$ si y sólo si para todo w_j, tal que w_jRw_i, $v(A,w_j)=1$

VII) $v(GA,w_i)=1$ si y sólo si para todo w_j, tal que w_iRw_j, $v(A,w_j)=1$.

Una fbf A es T-satisfacible si y sólo si existe un modelo-T, $<W,R,v>$, y un momento $w_i \in W$, tales que $v(A,w_i)=1$. Una fbf A es T-válida si y sólo si para todo modelo-T, $<W,R,v>$, y todo $w_i \in W$, $v(A,w_i)=1$.

Las propiedades de R, de las que se hablaba en ii), son las que van a quedar recogidas en los axiomas del sistema. En los sistemas básicos de tiempo lineal, R es transitiva. Si el tiempo es, por ejemplo, circular, R ha de ser reflexiva, simétrica y transitiva. Tal simetría no es habitual en la lógica temporal, puesto que la primera propiedad que se suele exigir a la relación de ulterioridad es la antisimetría.

IV. EL SISTEMA MÍNIMO

El sistema mínimo de lógica temporal, kt (Lemmon, 1965), está formado por los siguientes axiomas:

Ax. 0. Un conjunto suficiente de axiomas para derivar todas las tautologías de la lógica proposicional clásica Ax. 1, Ax. 2, Ax. 3, Ax. 4, Ax. 5 y Ax. 6.

Como regla de derivación tenemos el *modus ponens* (MP).

El lenguaje formal de kt consta de:

i) un conjunto enumerable de variables proposicionales, p, q, r, etc.
ii) $\rightarrow$ y $\neg$ como conectivas primitivas,
iii) (,) como signos auxiliares,
iv) A, B, etc. como variables metalingüísticas, y
v) F, P, G y H como operadores temporales.

Una fbf en kt será una concatenación de signos primitivos de kt, de alguno de los tipos siguientes:

1) Toda variable proposicional es fbf.
2) $\neg A$, donde A es fbf
3) $A \rightarrow B$, $(A \rightarrow B)$, donde A y B son fbfs
4) FA, donde A es fbf
5) PA, donde A es fbf
6) GA, donde A es fbf
7) HA, donde A es fbf.

V. OTROS SISTEMAS

El sistema para el tiempo lineal, en su primera formulación debido a Cocchiarella en 1965, añadiría a kt:

* En el lenguaje formal,
  ii) →, ¬, & y v, como conectivas primitivas,
  8) AvB, (AvB), donde A y B son fbfs, y
  9) A&B, (A&B), donde A y B son fbfs.
* En la axiomática, Ax. 7, Ax. 8 y Ax. 9.
* En la semántica, la relación R ha de cumplir las propiedades relativas a los axiomas 7, 8 y 9. Éstas son: transitividad (Ax. 7-P1), linearidad a la derecha (Ax. 8-P2) y linearidad a la izquierda (Ax. 9-P3),
  P1. (x)(y)(z)((xRy&yRz)→xRz)
  P2. (x)(y)(z)((xRy&xRz)→((x=y)vyRzvzRy))
  P3. (x)(y)(z)((yRx&zRx)→((x=y)vzRyvyRz))

El sistema para el tiempo ramificado tiene muchas formulaciones. El sistema básico ramificado es también debido a Cocchiarella y diferirá de kt sólo en que incluiría el Ax. 7 y la P1. Si el sistema es ramificado sólo hacia el futuro, pero lineal en el pasado (como el presentado por Rescher y Urquhart en 1971, llamado kb), añadiría el Ax. 10.

VI. EL SISTEMA OT

A partir de kb, se construyó un sistema llamado Ot (Okhamist Tense), expuesto por primera vez por Prior, que, siendo lineal en el pasado, interpreta en el futuro la posibilidad y la necesidad como ramificadas, mientras que F y G son lineales. Si se construye un modelo ramificado considerándolo como una colección de modelos lineales parciales desembocaría en una concepción lineal del tiempo. Pero el sistema Ot prevé esto, eligiendo arbitrariamente una ramificación para ser el futuro *prima-facie* del actual momento histórico. Esto permite hablar del futuro con pleno sentido, pues se está haciendo referencia a lo que hemos tomado como futuro actual y, al mismo tiempo, las ramificaciones permiten seguir hablando de posibilidades y necesidades. Para ello, en el lenguaje formal, se añaden L y M como operadores modales.

Como axiomas, Ot, tal y como es presentado en McArthur (1976), tiene Ax. 0, Ax. 1, Ax. 2, Ax. 3, Ax. 4, Ax. 5, Ax. 6, Ax. 7, Ax. 8, Ax. 9, Ax. 10, Ax. 11 y:

Ax. 15. L(A→B)→(LA→LB)

Ax. 16. MMA→MA

Ax. 17. LA, cuando A es una tautología de la lógica proposicional clásica

Ax. 18. LA→GA

Ax. 19. A→LPA, donde A no tiene ninguna ocurrencia de F.

Como regla de derivación tenemos el *modus ponens* (MP).

Esta axiomatización del sistema ockhamista no representa adecuadamente un tiempo ramificado en el futuro. Para una axiomatización de este tipo de tiempo puede verse Anderau, Enjalbert y Fariñas (1989), Zanardo (1985) y Gabbay (1994).

VII. CONSISTENCIA Y COMPLETUD

A continuación vamos a demostrar la consistencia y la completud para el sistema mínimo kt, axiomatizado en el apartado IV, así como algunas otras propiedades semánticas de interés.

El sistema kt será consistente, para un conjunto de modelos, si toda tesis de kt es T-válida en ese conjunto de modelos y kt será completo, para un conjunto de modelos, si toda fórmula T-válida es ese conjunto de modelos es una tesis de kt o, de manera equivalente, si toda fórmula consistente con kt es T-satisfacible en ese conjunto de modelos.

Teorema 1 (teorema de consistencia): kt es consistente para el conjunto de todos sus modelos

Prueba: Hay que mostrar que cada una de las tesis de kt es válida sobre el conjunto de sus modelos. Para ello basta con mostrar que cada uno de los axiomas de kt es válido.

Para demostrar que el Ax. 1 es válido, debemos mostrar que para todo $<W,R,v>$ y para cualquier w_i, si $v(G(A \rightarrow B),w_i)=1$ y $v(GA, w_i)=1$, entonces $v(GB, w_i)=1$. Por la hipótesis tenemos que siempre que w_iRw_j y $v(A,w_j)=1$, entonces $v(B,w_j)=1$ y que siempre que w_iRw_j, $v(A,w_j)=1$. De aquí se sigue inmediatamente que siempre que w_iRw_j, $v(B,w_j)=1$. La demostración del Ax. 2 es similar.

Para demostrar el Ax. 4, debemos mostrar que para todo $<W,R,v>$ y para cualquier w_i, si $v(A,w_i)=1$, entonces $v(GPA,w_i)=1$. Puesto que la conclusión es que para todo w_j, tal que w_iRw_j, hay un w_k, tal que w_kRw_j, y que $v(A,w_k)=1$, basta con que $w_k=w_i$. De forma similar, se demuestra el Ax. 3.

El Ax. 5 y el Ax. 6 se demuestran de manera inmediata. Puesto que para todo $<W,R,v>$ y w tenemos que $v(A,w)=1$, entonces $v(HA,w)=1$ y $v(GA,w)=1$.

Para la prueba de completud, tipo Henkin, se seguirá, fundamentalmente, Burgess (1984). El concepto central es el de conjunto máximamente consistente y, para llegar a él, necesitamos varias definiciones previas.

Def. 1. Γ es un conjunto *inconsistente* si y sólo si $\Gamma \vdash A \& \neg A$ para alguna fbf A.

Def. 2. Γ es un conjunto *consistente* si y sólo si Γ no es inconsistente.
Def. 3. Γ es un *conjunto máximamente consistente, CMC,* si y sólo si:
i) Γ es consistente, y
ii) si $\Gamma \cup \{A\}$ es consistente, $A \in \Gamma$.

Teorema 2 (lema de Lindenbaum): Si π es un conjunto consistente, hay un conjunto máximamente consistente Γ tal que $\pi \subseteq \Gamma$

Prueba:
Sea $A_1, \ldots, A_n$ una enumeración de las fbfs. Defínase la serie de conjuntos $\Gamma_0, \Gamma_1, \ldots$
i) $\Gamma_0 = \pi$
ii) Para cada $i \geq 0$, $\Gamma_{i+1} = \Gamma_i \cup \{A_i\}$ si $\Gamma_i \cup \{A_i\}$ es consistente. Si no es consistente $\Gamma_{i+1} = \Gamma_i$.
iii) $\Gamma = \cup_{0 \leq i < \infty} \Gamma_i$.
Es claro que existe un Γ tal que $\pi \subseteq \Gamma$:
a) Γ es consistente. Por inducción, Γ_0 es consistente y, por construcción, cada Γ_i lo es. Por tanto, $U\Gamma_i = \Gamma$ es consistente.
b) Si $\Gamma U\{A\}$ es consistente, $A \in \Gamma$. Si $\Gamma U\{A\}$ es consistente, hay un Γ_k tal que $\Gamma_k = \Gamma_{k-1} U\{A\}$.
Entonces $A \in \Gamma_k$ y, por tanto, $A \in \Gamma$.

Teorema 3: Si Γ es consistente y $\Gamma \vdash A$, $\Gamma U\{A\}$ es consistente

Prueba:
Si $\Gamma U\{A\}$ fuera inconsistente, tendríamos que $\Gamma U\{A\} \vdash B \& \neg B$, y, por lógica de proposiciones, que $\Gamma \vdash \neg A$, con lo cual $\Gamma \vdash A \& \neg A$ y Γ es inconsistente, lo que contradice la hipótesis.

Teorema 4: Si Γ es CMC,

i) $\Gamma \vdash A$ si y sólo si $A \in \Gamma$
ii) $A \& B \in \Gamma$ si y sólo si $A \in \Gamma$ y $B \in \Gamma$
iii) $A \vee B \in \Gamma$ si y sólo si $A \in \Gamma$ o $B \in \Gamma$
iv) $A \rightarrow B \in \Gamma$ si y sólo si $A \notin \Gamma$ o $B \in \Gamma$
v) $A \in \Gamma$ ó $\neg A \in \Gamma$.

Pruebas:
i) a) $A \in \Gamma$, entonces $\Gamma \vdash A$.
b) $\Gamma \vdash A$, Γ es consistente, $\Gamma U\{A\}$ es consistente. Entonces, $A \in \Gamma$ por def. 3.
ii) a) $A \& B \in \Gamma$, $\Gamma \vdash A \& B$, $\Gamma \vdash A$, $\Gamma \vdash B$, $A \in \Gamma$, $B \in \Gamma$.
b) $A \in \Gamma$ y $B \in \Gamma$, $\Gamma \vdash A$, $\Gamma \vdash B$, $\Gamma \vdash A \& B$, $A \& B \in \Gamma$.
iii) a) $AUB \in \Gamma$,
si $A \notin \Gamma$ y $B \notin \Gamma$, $\Gamma U\{A\}$ es inconsistente, $\Gamma U\{B\}$ es inconsistente, $\Gamma U\{A\} \vdash C \& \neg C$, $\Gamma U\{B\} \vdash D \& \neg D$, $\Gamma \vdash \neg A$, $\Gamma \vdash \neg B$, $\Gamma \vdash$

$\neg$A&$\neg$B, $\Gamma\vdash\neg$(AvB). Pero como AUB$\in\Gamma$, $\Gamma\vdash$AUB. Por tanto, Γ sería inconsistente.
b) A$\in\Gamma$ o B$\in\Gamma$. Por un lado, si A$\in\Gamma$, $\Gamma\vdash$A, $\Gamma\vdash$AvB, AvB$\in\Gamma$. Por otro lado, si B$\in\Gamma$, $\Gamma\vdash$B, $\Gamma\vdash$AvB, AvB$\in\Gamma$. Con lo cual AvB$\in\Gamma$.

iv) a) A$\rightarrow$B$\in\Gamma$, si A$\in\Gamma$ y B$\notin\Gamma$, $\Gamma\vdash$A$\rightarrow$B, $\Gamma\vdash$A, $\Gamma\vdash$B, B$\in\Gamma$. Entonces B$\in\Gamma$ y B$\notin\Gamma$.
b) A$\notin\Gamma$ o B$\in\Gamma$. Por un lado, si A$\notin\Gamma$, ΓU{A}$\vdash$C&$\neg$C, $\Gamma\vdash$ $\neg$A, $\Gamma\vdash\neg$A$\rightarrow$(A$\rightarrow$B), $\Gamma\vdash$ A$\rightarrow$B, A$\rightarrow$B$\in\Gamma$.
Por otro lado, si B$\in\Gamma$, $\Gamma\vdash$B, $\Gamma\vdash$B$\rightarrow$(A$\rightarrow$B), $\Gamma\vdash$A$\rightarrow$B, A$\rightarrow$B$\in\Gamma$. Por tanto, A$\rightarrow$B$\in\Gamma$.

v) Si A$\notin\Gamma$ y $\neg$A$\notin\Gamma$, ΓU{A}$\vdash$B&$\neg$B, ΓU{$\neg$A}$\vdash$C&$\neg$C, $\Gamma\vdash\neg$A, $\Gamma\vdash\neg\neg$A, $\Gamma\vdash\neg$A&$\neg\neg$A. Pero Γ es CMC, $\neg$A&$\neg\neg$A es inconsistente, luego A$\in\Gamma$ o $\neg$A$\in\Gamma$.

Teorema 5: Si tenemos dos CMs Γ y Γ', las cuatro cláusulas siguientes se cumplen, siendo equivalentes entre sí:

i) si A$\in\Gamma$, entonces PA$\in\Gamma$'
ii) si B$\in\Gamma$', entonces FB$\in\Gamma$
iii) si GC$\in\Gamma$, entonces C$\in\Gamma$'
iv) si HD$\in\Gamma$', entonces D$\in\Gamma$.

Prueba: Para mostrar que i) implica iii), se asume i) y que GC$\in\Gamma$. Entonces PGC$\in\Gamma$' y, por lógica de proposiciones y el Ax. 3, tenemos que PGC$\rightarrow$C, con lo que C$\in\Gamma$'. Se utilizan pruebas similares para los demás casos (que iii) implica ii), que ii) implica iv) y que iv) implica i)).

Def. 4. Para los CMCs Γ y Γ', decimos que Γ *es seguido en potencia por* Γ*'*, y escribimos A►B, si cumple las condiciones impuestas en las cláusulas del teorema 5. Intuitivamente, esto significaría que una situación del tipo descrito por Γ podría ser seguida de una situación del tipo descrito por Γ'.

Teorema 6: Si Γ es CMC,

i) y FA$\in\Gamma$, hay un CMC Γ' tal que Γ►Γ' y A$\in\Gamma$',
ii) y PA$\in\Gamma$, hay un CMC Γ'' tal que Γ''►Γ y A$\in\Gamma$''.

Prueba: i) El esquema de demostración sería como sigue. Tenemos FA$\in\Gamma$ y tenemos que conseguir Γ' tal que Γ►Γ' y A$\in\Gamma$'. Por def. 4 y teorema 5 i), si Γ►Γ', entontes para cualquier B$\in\Gamma$', se dé que PB$\in\Gamma$' y tal que A$\in\Gamma$', siendo Γ' un CMC. Por teorema 2, si demostramos que el conjunto Γ'_0 = {PB, A} es consistente, habríamos demostrado que puede conseguirse. Γ' sería simplemente una extensión suya. Tenemos ahora que demostrar que Γ'_0 es consistente. Habría que demostrar que para cualquier B$\in\Gamma$, PB&A es consistente. Para esto llega con demostrar que F(PB&A) es consistente (puesto que si fuese verdad que $\neg$(PB&A) ten-

dríamos $\neg$ F(PB&A), a partir de Ax. 5 y definición de G), lo que es cierto puesto que F(PB&A)$\in\Gamma$, ya que (B&FA)$\rightarrow$F(PB&A) es una fórmula demostrable y B&FA lo tenemos por hipótesis. La prueba para ii) seguiría un esquema similar.

Def. 5. Una *crónica* en un marco <W,R> (W y R entendidos de la forma habitual) es una función T que asigna a cada $w_i \in W$ un CMC $T(w_i)$. (Empleamos el concepto de marco en el sentido de Segerberg, tal como es recogido por Hughes y Cresswell, 1984). Intuitivamente, si W representa el conjunto de momentos y R la relación de ulterioridad, T proporcionaría una descripción completa de lo que pasa en cada momento. T es *coherente* si tenemos $T(w_i)T(w_j)$ siempre que w_iRw_j. Si T es coherente y satisface la condición i) siguiente es *profética,* si satisface la ii) es histórica:

i) Siempre que $FA \in T(w_i)$ hay un w_j tal que w_iRw_j y $A \in T(w_j)$.
ii) Siempre que $PA \in T(w_i)$ hay un w_j tal que w_jRw_i y $A \in T(w_j)$.
T es *perfecta* si es tanto histórica como profética.
Además, T será coherente si y sólo si satisface las dos condiciones siguientes:
iii) Siempre que $GA \in T(w_i)$ y w_iRw_j, entonces $A \in T(w_j)$.
iv) Siempre que $HA \in T(w_i)$ y w_jRw_i, entonces $A \in T(w_j)$.
Respecto a la evaluación, si v es la evaluación en <W,R> la *crónica inducida* Tv, en W_i, $Tv(w_i)$ estaría formada por todo A, tal que $v(A,w_i)=1$. Tv es siempre perfecta. Si T es una crónica perfecta en <W,R>, la *evaluación inducida* Vt, en w_i, $Vt(A,w_i)=1$ si y sólo si $A \in T(w_i)$.

Para probar la completud de kt para un conjunto de modelos, debemos mostrar que toda fórmula consistente A_0 es satisfacible. El teorema 7 ofrecerá una estrategia para mostrar que A_0 es satisfacible. Esta estrategia consiste en construir una crónica perfecta T, de acuerdo con la def. 5, en algún marco <W,R> que contenga un w_0 tal que $A \in T(w_0)$.

Teorema 7: Si T es una crónica perfecta en un marco <W,R> y v = Vt es la evaluación inducida por T, entonces T = Tv es la crónica inducida por v. En otras palabras tenemos que $v(A,w_i)=1$ si y sólo si $A \in T(w_i)$. En particular, cualquier miembro de cualquier T(w) es satisfacible en <W,R>

Prueba: Se demuestra por inducción sobre la complejidad de A. Por ejemplo para G, asumimos que $v(A,w_i)=1$ si y sólo si $A \in T(w_i)$ para A y lo probamos para GA. Por un lado, si $GA \in T(w_i)$, entonces por def. 5 (iii) siempre que haya un w_j tal que w_iRw_j, entonces $A \in T(w_j)$ y, por la hipótesis de la inducción, $v(A,w_j)=1$. Esto muestra que $v(GA,w_i)=1$. Por el otro lado, si $GA \notin T(w_i)$, entonces, puesto que $F\neg A =_{def} \neg GA$, $F\neg A \in T(w_i)$ y, por def. 5 (i), para algún w_j tal que w_iRw_j, tenemos que

$\neg A \in T(w_j)$ y $A \notin T(w_j)$. De donde, por hipótesis de la inducción, $v(A,w_j)=0$. Esto muestra que $v(GA,w_i)=0$.

Def. 6. Tómese un conjunto infinito enumerable X. Sea M el conjunto de todos los triples $<W,R,T>$, tales que:

i) $W \neq \emptyset$ y $W \subseteq X$,
ii) $R \subseteq W^2$ y es antisimétrica.
iii) T es una crónica coherente en $<W,R>$.
Para $\mu = <W,R,T>$ y $\mu' = <W',R',T'>$ en M, decimos que μ' es una *extensión* de μ si, estando las relaciones y funciones identificadas con conjuntos de pares ordenados, tenemos:
i') $W \subseteq W'$.
ii') $R = R' \cap (W x W')$,
iii') $T \subseteq T'$.
Una condición de la forma de la establecida por la def. 5 (i) o (ii) será llamada *no nacida* para $\mu = <W,R,T>$ si su antecedente no se cumple, es decir, si $w_i \notin W$ o si $w_i \in W$ pero FA o PA, dependiendo del caso, no pertenece a $T(w_i)$. Será llamada *viva* para μ si su antecedente se cumple pero su consecuente no; en otras palabras, no hay ningún $w_i \in W$ tal que $w_i R w_j$ o $w_j R w_i$, según el caso, y $A \in T(w_j)$. Será llamada *muerta* para μ si el consecuente se cumple.

Teorema 8: Si $\mu = <W,R,T>$, para cualquier condición de forma de la def. 5 (i) o (ii) que esté viva para μ, existe una extensión $\mu' = <W',R',T'>$ de μ en la que esta condición esté muerta

Prueba: Tomemos la condición de la forma de la def. 5 (i). Si $w_i \in W$ y $FA \in T(w_i)$, por el teorema 6 (i), hay un CMC Y, tal que $T(w_i) \blacktriangleright Y$ y $A \in Y$. Esto sirve para fijar un w_j, tal que $w_j \in X\text{-}W$, y para construir los conjuntos:
a) $W' = W \cup \{w_j\}$,
b) $R' = R \cup \{<w_i, w_j>\}$ y
c) $T' = T \cup \{<w_j, Y>\}$.

Teorema 9 (teorema de completud): kt es completo para un conjunto de modelos Σ

Prueba: Dada una fórmula consistente A_0, deseamos construir un marco $<W,R>$ y una crónica perfecta T en él, con $A_0 \in T(w_0)$ para algún w_0. Para este fin, fijamos una enumeración $w_0, w_1, w_2,...$ de X y una enumeración $A_0, A_1, A_2,...$ de todas las fórmulas de kt. A la condición de la def. 5 (i) (o (ii)) se le asigna un número de código, dependiente de esta enumeración. Se fija un CMC Γ_0, con $A_0 \in \Gamma_0$, y sea $\mu_0 = <W_0, R_0, T_0>$, donde $W_0 = \{w_0\}$, $R_0 = \emptyset$ y $T_0 = \{<w_0, \Gamma_0>\}$. Si μ_n está definido, con-

sidérese aquella condición que, entre todas las que estén vivas para μ_n tenga el número de código menor. Sea μ_{n+1}, según el teorema 8, una extensión de μ_n para la cual la condición esté muerta. Sea $<W,R,T>$ la unión de los $\mu_n = <W_n, R_n, T_n>$; de manera más precisa, sea W la unión de los W_n, R la unión de los R_n y T la de los T_n. Así se verifica que T es una crónica perfecta de $<W,R>$.

BIBLIOGRAFÍA

Auderau, E., Enjalbert, P. y Fariñas del Cerro, L. (1989), *Logique temporelle, semantique et validation de programmes parallèles,* Masson, Paris.

Bahsoun, J. P. (1988), *Expression de la synchronisation dans un module control paré priorité: implantation et méthode de preuve,* Université Paul Sabatier, Toulouse.

Bakker, J. W., Roever, W. P y Rozemberg, G. (eds.) (1989), *Linear Time, Branching Time and Partial Order in Logics and Models for Concurrency,* Springer Verlag, Heidelberg.

Banieqbal, B. y Barringer, H. (1986), *A Study of an Extended Temporal Language and a Temporal Fixed Point Calculus,* Technical report UMCS-86-10-2, Department of Computer Science, University of Manchester.

Barringer, H., Fisher, M., Gabbay, D., Gouch, G. y Owens, R. (1986), «METATEM: A framework for programming in temporal logic»: *Stepwise Refinement of Distributed Systems.*

Bras, M. (1990), *Calcul des structures temporelles du discours,* Université Paul Sabatier, Toulouse.

Burgess, J. P. (1979), «Logic and Time»: *Journal of Symbolic Logic, 44,* 566-582.

Burgess, J. P. (1982), «Axioms for tense logic»: *Notre Dame Journal of Formal Logic, 23,* 367-383.

Burgess, J. P. (1984), «Basic tense logic», en Gabbay y Guenthner (eds.), 1984.

Cocchiarella, N. (1965), *Tense and modal logic: a study of the topology of temporal reference,* tesis doctoral, Universidad de California en Los Angeles (UCLA).

Cocchiarella, N. (1984), «Philosophical perspectives on quantification in tense and modal logic», en Gabbay y Guenthner (eds.), 1984.

Díaz, J. (ed.) (1983), *Automates, Languages and Programming,* Springer Verlag, Heidelberg.

Dummet y Lemmon (1959), «Modal logics between S4 and S5»: *Zeitschrift für matematische Logik und Grundlagen der Mathematik,* vol. 3, VEB Deutsches Verlag der Wissenschaften, Berlin.

Emerson, E. A. y Srinivisan, J. (1989), «Branching Time Temporal Logic», en Bakker, Roever y Rozemberg (eds.), 1989.

Gabbay, D. (1975), «Model theory for tense logic and decidibility results for non classical logics»: *Ann. Mathematical Logic, 8,* 185-295.

Gabbay, D. y Guenthner, F. (eds.) (1984), *Handbook of Philosophical Logic,* Reidel, Dordrecht.

Gabbay, D. M., Hodkinson, I. y Reynolds, M. (1994), *Temporal Logic. Mathematical Foundations and Computational Aspects,* vol. 1, Clarendon Press, Oxford.

Goldblatt, R. (1978), «Diodorean Modality in Minkowski Spacetime»: *Studia Logica, 39.*

Goldblatt, R. (1992), *Logics of Time and Computation,* segunda edición revisada y ampliada, CSLI Lecture Notes, Stanford.

Haack, S. (1974), *Deviant Logic,* Cambridge University Press.

Haack, S. (1978), *Philosophy of Logics,* Cambridge University Press.

Hughes, G. y Cresswell, H. (1968), *An Introduction to Modal Logic,* Methnen and Co.

Hughes, G. (1984), *A Companion to Modal Logic,* Methnen and Co.
Kamp, H. (1968), *On Tense Logic and the Theory of Order,* Universidad de California en Los Angeles (UCLA).
Kowalski, R. y Sergot, M. (1986), «A Logic-based Calculus of Events»: *New Generation Computing, 4,* 67-95.
León, J. C. (1993), «Estructuras semánticas para la lógica del tiempo indeterminista»: *Actas del I Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España,* UNED, Madrid.
McArthur, R. P. (1976), *Tense Logic,* Reidel, Dordrecht.
Makim, M. y Davis, C. (1976), «Temporal Modalities and the Future»: *Notre Dame Journal of Formal Logic,* abril 1976.
Manna, Z. y Pnueli, A. (1983), «Proving precedence properties: the temporal way», en Díaz (ed.), 1983.
Mates, B. (1952), *Stoic Logic,* University of California Press, Los Angeles.
Nishimura, H. (1979), «Is the semantics of branching structures adequate for non-metric ockhamist tense logics?»: *Journal of Philosophical Logic, 8.*
Peña, L. (1989), «Algunos desarrollos recientes en la articulación de lógicas temporales», en *Actas del IV Congreso de lenguajes naturales y lenguajes formales,* Universidad de Barcelona.
Pnueli, A. (1977), «The temporal logic of programs»: *Proc. 18th Ann. Symp. Foundations of Computer Science,* RI, New York, IEEE, 46-57.
Prior, A. N. (1957), *Time and Modality,* Clarendon Press, Oxford.
Prior, A. N. (1967), *Past, Present and Future,* Clarendon Press, Oxford.
Prior, A. N. (1968), *Papers on Time and Tense,* Clarendon Press, Oxford.
Quine, W. O. (1960), *Word and Object,* MIT Press, Cambridge, Mass.
Rescher, N. y Urquhart, A. (1971), *Temporal Logic,* Springer Verlag, Berlin.
Segerberg, K. (1970), «Model logics with linear alternative relations»: *Theoria, 36.*
Shoham, Y. (1988), *Reasoning about change. Time and causation from the standpoint of artificial intelligence,* MIT Press, Cambridge, Massachussetts.
Strawson, P. F. (1952), *Introduction to Logical Theory,* Barnes and Noble.
Thayse, A. y otros (1989), *Approche Logique de l'Intelligence Artificielle. 2. De la logique modale à la logique de bases de données,* Bordas, Paris.
Thomason, R. H. (1970), «Indeterministic time and truth value gaps»: *Theoria, 36.*
Thomason, R. H. (1972), «Semantic Analysis of Tense Logic»: *The Journal of Symbolic Logic, 37.*
Thomason, R. H. (1984), «Combinations of tense and modality», en Gabbay y Guenthner (eds.), 1984.
Van Benthem, J. F. A. K. (1978), «Tense logic and standard logic»: *Logique et analyse, 80,* 47-83.
Van Benthem, J. F. A. K. (1983), *The Logic of Time,* Reidel, Dordrecht.
Wolper, P. (1982), *Synthesis of Communicating Processes from Temporal Logic Specifications,* Stanford University.
Zanardo, A. (1985), «A finite axiomatization of the set of strongly valid Ockamist formulas»: *Journal of Philosophical Logic, 14,* 447-468.

LA LÓGICA CUÁNTICA

Sergio F. Martínez Muñoz

I. INTRODUCCIÓN

La mecánica cuántica constituye la teoría más revolucionaria y fundamental de la física moderna. Si bien su adecuación empírica (en tanto que descripción de procesos estadísticos) está ampliamente confirmada, su estructura conceptual nunca ha sido elucidada a cabalidad. Esta estructura conceptual es tan diferente de la estructura conceptual de las teorías de la física clásica que desde los inicios de la teoría ha sugerido propuestas radicales en los fundamentos de la teoría de la probabilidad, la lógica, y la filosofía en general. En este artículo veremos cómo el examen de la estructura semántica de la teoría cuántica nos lleva al estudio de una clase de *lógicas no clásicas* genéricamente conocidas con el nombre de lógicas cuánticas. No asumiremos ningún conocimiento de la física cuántica, pero sí cierta familiaridad con el análisis semántico de la lógica (ver el artículo 1 en este volumen). En la segunda sección muestro cómo surge la lógica cuántica a partir de la consolidación de una analogía entre la estructura semántica de la teoría cuántica y la estructura semántico-algebraica del cálculo proposicional clásico. En la tercera sección examino muy brevemente intentos, sobre todo interesantes desde un punto de vista histórico, de entender la lógica cuántica como una lógica de varios valores. En la cuarta sección quiero dar una idea de cómo puede entenderse la lógica cuántica como lógica formal, reconstruida sintácticamente. En la quinta sección resumo algunos resultados sobresalientes de los esfuerzos por encajar a la lógica cuántica dentro de la semántica de marcos de Kripke, y de su interpretación como lógica modal. En el apéndice matemático se incluyen las definiciones básicas de la teoría de retículos necesarias para darle una precisión mínima a nuestra presentación.

II. LA LÓGICA CUÁNTICA COMO ESTRUCTURA PROPOSICIONAL NO CLÁSICA

La lógica cuántica fue propuesta inicialmente por Garett Birkhoff y John von Neumann (1936), como un intento por dar una solución radical al problema de la interpretación de la teoría cuántica. Su objetivo explícito era descubrir la estructura lógica que yace debajo de las teorías físicas que como la mecánica cuántica no se conforman a la lógica clásica. Sugirieron ellos en este trabajo seminal que la transición de la mecánica clásica a la mecánica cuántica involucra el paso de un cálculo proposicional clásico a un cálculo proposicional con una estructura no clásica. Sucintamente, la tesis de Birkhoff y von Neumann era que deberíamos considerar a una cierta estructura algebraica generada por la teoría cuántica como el álgebra de Lindebaum-Tarski de una nueva lógica, la lógica del mundo empírico, asumiendo que la mecánica cuántica es la teoría física que describe más fielmente ese mundo empírico.

Intentos por clarificar y elaborar esta propuesta desde perspectivas muy diferentes han generado una serie de investigaciones muy variadas. Principio por clarificar el sentido en el que la transición entre la física clásica y la física cuántica sugiere un cambio de lógica. Esto requiere que establezcamos una relación entre la estructura semántica de teorías físicas y un análisis semántico de la lógica. Esto lo haremos partiendo de la concepción semántica de teorías físicas que se origina con los trabajos de Beth y que ha sido desarrollada posteriormente por Van Fraassen (y otros). Llamaremos *lógica (proposicional) concreta* a una lógica (proposicional) que describe las relaciones semánticas entre las sentencias elementales en las que se ha fijado de antemano una intensión fija para los términos predicacionales (fija en el sentido que es respetada por todas las valuaciones admisibles). En este caso hablamos de un *lenguaje semi-interpretado* (Van Fraassen, 1970). Tradicionalmente la lógica se concibe como caracterizando la validez en virtud de la forma de los argumentos *únicamente;* a la lógica así entendida la llamaremos *lógica formal.* En lógicas concretas, a diferencia de las lógicas formales, se utilizan criterios semánticos, además de los puramente formales, para juzgar la validez de argumentos. Estos criterios semánticos adicionales, en el caso de la lógica cuántica concreta, se consideran dados implícitamente por la teoría cuántica. En esta sección hablamos de lógica cuántica concreta siempre. Posteriormente diremos algo muy breve acerca de la lógica cuántica formal. Una última aclaración previa es que hablaremos de la lógica cuántica haciendo referencia únicamente al cálculo proposicional cuántico. Esto no es una distorsión seria de nuestra presentación ya que las características peculiares de la lógica cuántica surgen al nivel del cálculo proposicional. El lector interesado en un compendio de la lógica cuántica de primer orden puede consultar Dalla Chiara (1986).

Según Van Fraassen, una teoría física puede caracterizarse por medio de un *lenguaje semi-interpretado* y un conjunto de leyes. El lenguaje semi-interpretado L, el portador de la estructura lógica (concreta) de la teo-

ría, consiste en una terna <E,H,h>; E es un conjunto de «sentencias elementales»[1]. Una sentencia elemental tiene la forma «La magnitud M tiene un valor en el conjunto (de Borel) X». La idea intuitiva es que M describe una propiedad física de un sistema dado. Por ejemplo, una magnitud de los sistemas conocidos como «partículas clásicas» es la velocidad V de una partícula. Una sentencia elemental es la sentencia «V tiene un valor en el intervalo [1,2]» (las unidades de la magnitud se dejan implícitas)[2]. H es el *conjunto de estados posibles* del sistema en cuestión. La función h es una *función de satisfacción* que asigna a cada sentencia elemental A en E el conjunto h(A) de estados que satisfacen A. A cada sentencia elemental A (un objeto sintáctico) corresponde la *proposición* h(A) (un objeto semántico). El conjunto de proposiciones elementales es la imagen h[E] de E bajo h.

Podemos ahora formular informalmente ciertas relaciones semánticas familiares en el contexto de la lógica cuántica:

1. A es *verdadera* si y sólo si el estado de un sistema se representa por un estado de h(A).
2. A es válida si y sólo si h(A) = H.
3. A es una *consecuencia semántica* de B si y sólo si $h(B) \subseteq h(A)$.

El *álgebra proposicional* de un lenguaje es el conjunto de proposiciones elementales h(E) junto con las operaciones lógicas asociadas con ese lenguaje. Los lenguajes (semi-interpretados) cuánticos tienen una estructura sintáctica pobre. Las sentencias son todas atómicas. La estructura lógica de un lenguaje cuántico es más bien una característica de su estructura semántica tal y como ésta se expresa a través de su álgebra proposicional. Esta estructura puede expresarse en términos de conectores definidos semánticamente. Aquí no podemos adentrarnos en la presentación detallada que requeriría una discusión a fondo del problema de la introducción de los conectores en la lógica cuántica. Daremos sin embargo una idea de la problemática involucrada y de las razones de su interés filosófico. Seguiremos la convención de identificar dos sentencias A y B cuando h(A) = h(B), esto es, cuando las sentencias son semánticamente equivalentes.

En las definiciones siguientes A, B, C, D son sentencias en un lenguaje L = <E,H,h>.

1. Hay una tendencia en la lógica cuántica, empezando con el trabajo de Birkhoff y von Neumann, a hablar indistintamente de propiedades de sistemas físicos y de proposiciones. Esta ambigüedad reaparece en la manera como hemos definido el conjunto de sentencias elementales. Más correctamente E es un conjunto de predicados monádicos elementales. Las sentencias elementales propiamente dichas pueden construirse a partir de estos predicados elementales en el contexto de una teoría física particular. Este tipo de ambigüedades las ignoraremos en pro de una mayor claridad expositiva.

2. Una magnitud puede definirse de manera abstracta como un conjunto de proposiciones (o propiedades) mutuamente excluyentes (*i.e.* tal que a lo más uno de los valores de la magnitud es el caso en un momento dado). Esta es la definición de magnitud apropiada en la formulación de una teoría de la mecánica en el marco de una teoría de retículos.

Definición: una sentencia C es la *conjunción* de A y B precisamente si $h(C) = h(A) \cap h(B)$.

Definición: una sentencia D es la *disyunción exclusiva* de A y B precisamente si $h(D) = h(A) \cup h(B)$.

Definición: una sentencia A en un lenguaje L es una *negación exclusiva* de la sentencia B precisamente cuando $h(A) = H - h(B)$.

Las operaciones «$\cap$», «$\cup$», y « – » son las operaciones usuales de la teoría de conjuntos (intersección, unión y complementación relativa). *Estas definiciones corresponden a los* conectores *clásicos (por lo general introducidos sintácticamente) si la estructura proposicional es clásica.* Esto es, si el conjunto de proposiciones es el conjunto P(H), el conjunto potencia del conjunto de los estados posibles al que llamamos H, entonces $\langle P(H), \cap, \cup \rangle$ es un álgebra booleana de conjuntos isomórfica al álgebra de Lindebaum-Tarski del cálculo proposicional clásico. Similarmente, podríamos definir otros conectores modales e intensionales en este marco conjuntista. Decimos que un lenguaje es cerrado con respecto a la conjunción o disyunción (exclusiva) si cada par de sentencias tiene una conjunción o disyunción (exclusiva). Decimos que un lenguaje es cerrado con respecto a la negación (exclusiva) si cada sentencia en L tiene una negación (exclusiva). Si la conjunción corresponde a la operación de intersección de conjuntos en P(H), y la disyunción corresponde a la unión de conjuntos en P(H), entonces es claro que las operaciones lógicas son cerradas en H (ya que por definición de conjunto potencia, P(H) incluye todos los conjuntos que puedan formarse por medio de las operaciones de conjuntos). Sin embargo, *si la estructura impuesta en H por la teoría es tal que el álgebra proposicional no corresponde al álgebra de conjuntos generada por P(H), entonces no está garantizado que las operaciones lógicas, correspondientes a los* conectores exclusivos *tal y como fueron definidas arriba, sean cerradas en H.* Por ejemplo, los lenguajes intuicionistas y las lógicas de varios valores no son cerrados con respecto a la negación exclusiva. Son cerrados con respecto a otro tipo de negación «selectiva» o de «alcance restringido», que es la negación semánticamente apropiada para estos lenguajes[3]. Este tipo de negación selectiva se caracteriza porque una sentencia A y su negación selectiva A^* pueden ser ambas no verdaderas simultáneamente (véase por ejemplo Rasiowa, 1974).

Algo similar sucede en la lógica cuántica. El conjunto H, según la mecánica cuántica, tiene asociada una estructura proposicional que no es cerrada con respecto a la negación exclusiva; además, y esto es algo peculiar de la lógica cuántica, no es cerrada con respecto a la disyunción exclusiva.

En la mecánica cuántica el conjunto de estados H tiene una estructura matemática significativamente diferente al conjunto de estados clá-

3. Van Fraassen le atribuye esta distinción entre «negación exclusiva» y «negación selectiva» *(choice negation)* a Mannoury, en sus trabajos de fundamentación del intuicionismo (ver Van Fraassen, 1974).

sicos (es un espacio complejo separable de Hilbert). Las sentencias elementales corresponden en este caso a subespacios (cerrados) del espacio H, y hay subconjuntos de H que no son subespacios. No es necesario que el lector entienda esta terminología matemática para captar la diferencia básica entre un lenguaje clásico y un lenguaje cuántico, ya que la diferencia, como vemos a continuación, se refleja en la semántica no-clásica de un lenguaje cuántico. Para explicar esta afirmación requeriremos el uso de algunos términos algebraicos que se definen en el apéndice.

El *álgebra proposicional cuántica* es una estructura (reticular) cerrada bajo la operación de intersección en H, pero no es cerrada ni bajo la unión de conjuntos, ni bajo la operación de complemento relativo (que corresponde a la negación exclusiva). No obstante, la lógica cuántica tiene definiciones alternativas de operaciones algebraicas cerradas que pueden pensarse como las operaciones lógicas correspondientes a la negación y disyunción cuántica. Estas operaciones son la *ortocomplementación* y la *operación de junta* (el resultado de la cual es el «extremo superior» de un par de elementos) en el retículo de los subespacios de H. No entraremos a definiciones detalladas, en su lugar pasamos a ilustrar estas operaciones en un ejemplo simple de un álgebra proposicional cuántica.

El conjunto de los subespacios de un espacio euclidiano de tres dimensiones, E, con las operaciones de intersección de conjuntos, $\cap$, y la operación de suma lineal $\oplus$, forma un retículo cuántico, esto es, un retículo que puede interpretarse como una lógica cuántica. Los subespacios de este espacio son el origen del sistema de coordenadas que se identifica con el cero del retículo, los subespacios de una dimensión (que geométricamente corresponden a las líneas que atraviesan el origen), los subespacios de dos dimensiones (que geométricamente corresponden a los planos que intersectan el origen en cualquier ángulo), y el espacio total, el único subespacio de tres dimensiones. La intersección de dos líneas cualesquiera es el origen, la suma lineal de dos líneas es el subespacio (plano) generado por las líneas. La intersección de dos planos es la línea en que se intersectan, y la suma lineal de dos planos (diferentes) es el espacio total H(S). Nótese que en este retículo $< E, \cap, \oplus >$ podemos definir siempre un orden parcial como sigue: $A \leq B$ si $A \oplus B = B$. Este orden parcial en nuestro ejemplo corresponde a la relación de inclusión de conjuntos. Una magnitud en este ejemplo es un conjunto de subespacios mutuamente excluyentes (*i.e.* para todo par de subespacios en la magnitud su intersección es $\emptyset$) y cuya unión es el conjunto total. Una magnitud máxima en general es un conjunto máximo de proposiciones mutuamente excluyentes. En nuestro ejemplo una magnitud máxima es un conjunto de tres líneas linealmente independientes (no paralelas entre sí) que geométricamente describen un sistema de coordenadas. Es fácil ver con un ejemplo que este retículo no es distributivo. Consideremos una línea D que no coincide con ninguna de las líneas (direcciones) A, B, C. Es claro que $D \wedge A = 0$, $D \wedge B = 0$, $D \wedge C = 0$, y por lo tanto: $(D \wedge A) \vee (D \wedge B) \vee (D \wedge C) = 0$. Sin embargo, $(A \wedge B \wedge C) \wedge D = D \neq 0$.

Este ejemplo muestra que un retículo proposicional cuántico no satisface la *ley distributiva*. Sin embargo, puede mostrarse que un retículo cuántico satisface la condición de *ortomodularidad:* $A < B \Rightarrow B = A \vee (B \wedge A^{\perp})$. $A^{\perp}$ es el orto-complemento de A, correspondiente a la negación selectiva de A (ver apéndice). Una buena parte del desarrollo del programa de lógica cuántica posterior al trabajo de Birkhoff y von Neumann puede verse como un intento por interpretar esta ley de orto-modularidad (ver Jauch, 1968, por ejemplo), y de entender el sentido en que, supuestamente, esta ley ortomodular podría jugar un papel similar al que juega la distributividad en los cálculos clásicos. En el trabajo original de Birkhoff y von Neumann, así como en el ejemplo que dimos anteriormente, la estructura proposicional satisfacía una condición más fuerte que la ortomodularidad, la modularidad. Un retículo es modular si satisface la siguiente ley: $A \leq B \Rightarrow (A, B, X)$ es una tripleta distributiva, para cualquier X en L. Puede mostrarse que un retículo modular es la estructura más débil en la cual una teoría de las probabilidades clásica puede formularse. Esto les sugirió a Birkhoff y von Neumann que la estructura proposicional podía pensarse como una teoría generalizada de conjuntos en la que la dimensionalidad jugaba el papel de la cardinalidad y las probabilidades de transición entre sucesos eran inducidas por automorfismos del retículo. Este programa sin embargo nunca llegó a desarrollarse más allá de algunas notas no publicadas de von Neumann, aunque sí dio lugar al desarrollo de una teoría matemática de envergadura (La teoría de las geometrías continuas de von Neumann).

Hay una serie de trabajos que parten de la convicción de que la estructura reticular incluye estructura que no puede justificarse físicamente, y tratan de estudiar estructuras más débiles. Trabajos significativos en esta dirección son los de Kochen y Specker (1967), Strauss (1937). Nuestra exposición, así como estos trabajos recién mencionados, se enmarcan en lo que se denomina el enfoque algebraico a la lógica cuántica.

La lógica cuántica es también otras cosas. Menciono a continuación algunas de las principales tradiciones alternativas.

III. LA LÓGICA CUÁNTICA COMO LÓGICA DE VARIOS VALORES

Una formulación de la lógica cuántica como una lógica de varios valores fue propuesta por Hans Reichenbach (basado en un formalismo de Lukasiewicz) en su libro sobre los fundamentos filosóficos de la mecánica cuántica (Reichenbach, 1944). Reichenbach proponía una lógica de tres valores como una manera de resolver los problemas de la interpretación de las descripciones mecánico-cuánticas del mundo. Sin embargo, a diferencia de la manera como presentamos la lógica cuántica en la sección anterior, Reichenbach pretendía no una descripción de la estructura lógica de los postulados descriptivos de la mecánica cuántica, sino más bien la formulación de la base *lógico-lingüística* para la formulación de una teo-

ría alternativa. La exposición más clara y detallada, así como el análisis más a fondo, de la propuesta de Reichenbach se encuentra en una serie de trabajos de Gary Hardegree (ver, por ejemplo, Hardegree, 1977). Hardegree introduce una distinción entre lenguaje observacional y lenguaje de la formulación de la teoría y arguye convincentemente que no hay necesidad de implementar una lógica no clásica en el lenguaje de formulación de la teoría, si bien reconoce que el lenguaje observacional tiene una estructura no clásica. El programa de Reichenbach ha sido abandonado, pero las intuiciones básicas de su enfoque se han retomado y desarrollado fértilmente en el enfoque modal del que hablaremos en la sección V.

IV. LA LÓGICA CUÁNTICA COMO LÓGICA FORMAL

La lógica cuántica concreta trata de la estructura lógica de las proposiciones generadas por la estructura semántica de la mecánica cuántica. La lógica cuántica formal, como toda lógica formal, requiere que hagamos explícita una sintaxis y que abandonemos los postulados semánticos (y el lenguaje semi-interpretado) que en la lógica cuántica concreta nos limitan a un discurso de objetos. En la lógica formal tanto los objetos como los símbolos que utilizamos para referirnos a ellos son parte de la teoría. No hay, sin embargo, una sola sintaxis que pueda reconstruirse a partir de la estructura lógico-algebraica de la teoría cuántica. Los diferentes sistemas que se han propuesto y se siguen proponiendo tienen diferencias lógicas y metalógicas importantes. Para varios sistemas de lógica cuántica formal se ha demostrado su *corrección* y *completitud*. Por lo general las pruebas de completitud proceden de la manera usual, construyendo el álgebra asociada de Lindenbaum-Tarski determinada por la axiomatización propuesta (ver, por ejemplo, Stachow, 1976; Dalla Chiara, 1986).

Con respecto a la decidibilidad de los diferentes sistemas de lógica cuántica hay una serie de resultados interesantes. La decidibilidad de la lógica clásica puede mostrarse utilizando tablas de verdad. Este método no funciona para demostrar la decidibilidad de la lógica intuicionista, pero en este (y muchos otros casos) la decidibilidad puede demostrarse a través del establecimiento de la propiedad del modelo finito (una técnica muy desarrollada en lógicas modales). Estas técnicas no pueden aplicarse en el caso de muchas lógicas cuánticas formales. Otras técnicas han sido ensayadas, pero para una buena parte de las lógicas cuánticas formales la decidibilidad no ha sido demostrada. En el caso de lógicas cuánticas débiles (orto-lógicas por ejemplo) es posible demostrar la decibilidad por medio de la traducción en lógicas modales. Pero para lógicas cuánticas que incluyen la propiedad ortomodular, y que supuestamente serían aquellas lógicas sancionadas por la mecánica cuántica como físicamente significativas, esta técnica no es aplicable. Goldblatt (1984) ha mostrado que la orto-modularidad del retículo de subespacios de un espa-

cio de Hilbert H no está determinada por ninguna propiedad de primer orden de la relación de ortogonalidad. Este resultado sugiere que la lógica cuántica tiene limitaciones de capacidad de expresión serias, por lo menos si se identifica como es usual con lógicas ortomodulares. Quizás valga la pena retomar la idea inicial de Birkhoff y von Neumann según la cual la lógica cuántica obedece la condición de modularidad, partiendo de la hipótesis de que la formulación de la mecánica cuántica en términos de espacios de Hilbert es sólo aproximadamente correcta. Lo más probable es que este tipo de desarrollo tenga que esperar nuevos adelantos en la manera de conceptualizar teorías físicas.

V. SEMÁNTICA DE KRIPKE E INTERPRETACIÓN MODAL DE LA LÓGICA CUANTICA

La semántica de Kripke desarrollada en los años cincuenta y sesenta de este siglo (por Kripke y otros) es una teoría unificada de la semántica que permite una clasificación bastante general de muchos sistemas lógicos (ver capítulo 12 en este volumen). Es posible extender esta teoría y formular una semántica de marcos de Kripke para la lógica cuántica. Una característica de los modelos de Kripke generalizados que resultan ser adecuados para la lógica cuántica es que la relación de acceso es reflexiva y simétrica, pero no transitiva. En las lógicas no-clásicas más comunes (como la lógica intuicionista y muchas lógicas modales) la relación es por lo menos reflexiva y transitiva. Es posible también dar una semántica algebraica para la lógica cuántica y mostrar que esta semántica es equivalente a la semántica de modelos de Kripke (Dalla Chiara, 1986).

No sólo es posible dar una semántica de Kripke para la lógica cuántica, sino que es también posible traducir la lógica cuántica a una lógica modal de una manera paralela a la traducción de Mckinsey-Tarski de la lógica intuicionista en la lógica modal S_4. Goldblatt (1974) ha hecho una traducción de una lógica cuántica débil, que él llama orto-lógica (en la que los marcos, llamados por él orto-marcos, son orto-retículos) en el sistema modal B. Posteriormente Dishkant (1977) ha construido una traducción de la lógica orto-modular en un sistema que él llama B+, que es intermedio entre B y S_5.

VI. RESUMEN Y CONCLUSIONES

En la primera parte de este artículo hemos visto cómo surgió la lógica cuántica a partir del desarrollo de una analogía entre la estructura reticular de la lógica clásica concreta (generada por la física clásica) y la estructura reticular de las proposiciones físicas sancionadas por la mecánica cuántica. La mecánica cuántica es supuestamente la teoría más confiable y general que tenemos. Esto sugiere que la lógica cuántica es *la*

lógica del mundo empírico. Putnam, elaborando ideas de Finkelstein, ha sugerido la analogía con el abandono de la geometría euclideana debido al desarrollo de la teoría de la relatividad de Einstein (Putnam, 1969; Finkelstein, 1969). Sin embargo, como hemos visto, no hay una sola lógica cuántica, sino una serie de sistemas lógicos con diferencias lógicas y meta-lógicas significativas. Además, en tanto que la relación de acceso de la lógica cuántica no reciba una interpretación física y lógica satisfactoria no es posible clarificar el sentido en que se propone la lógica cuántica como lógica alternativa, o por lo menos no es clara la pertinencia de la adecuación empírica de la mecánica cuántica para tal proyecto.

VII. APÉNDICE

En este apéndice se incluyen algunas definiciones de la teoría de retículos requeridas para clarificar la exposición.

Un conjunto ordenado es un par ordenado $L = \langle A, \leq \rangle$, donde A es un conjunto no vacío y $\leq$ es una relación parcialmente ordenada. L es un retículo si además para cada par de elementos existen el extremo superior y el extremo inferior. Un elemento z es *el extremo inferior* de un par de elementos $\{x,y\}$ si $x \geq z$, $y \geq z$, y además, si hay otro elemento w con la misma propiedad, entonces $z \geq w$. De manera dual se define la noción de extremo superior. El extremo inferior de un par de elementos se designa por el símbolo $x \wedge y$, y se lee «cuña de x, y». El extremo superior de un par de elementos se designa por el símbolo $x \vee y$, y se lee «junta de x, y». El ejemplo paradigmático de un retículo es el retículo formado por todos los subconjuntos de un conjunto (el conjunto potencia) con las operaciones de intersección y unión.

Un retículo ortocomplementado, o simplemente, un *orto-retículo,* es un retículo con 0 y 1 y con una ortocomplementación. Una *ortocomplementación* es un mapeo de L en L que satisface:

i) $a \vee a^{\perp} = 1$, $a \wedge a^{\perp} = 0$
ii) $a \leq b \Rightarrow a^{\perp} \geq b^{\perp}$
iii) $a^{\perp\perp} = a$

Una tripleta de elementos es *distributiva* si

(D) $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$

Un retículo es *distributivo* si para cada tripleta a, b, c en L se satisface (D). Un retículo es *modular* si satisface la condición

(M) $a \leq b \Rightarrow (a,b,x)$

es una tripleta distributiva para todo $x \in L$. Un retículo es *orto-modular* si satisface la siguiente condición:

(OM) $a \leq b \Rightarrow b = a \vee (b \wedge a^{\perp})$ para cualquier par de elementos a y b.

Un retículo booleano es un orto-retículo distributivo. Para retículos distributivos la ortocomplementación es un automorfismo dual, por lo que el orto-complemento de un elemento es único. El retículo $\langle P(A),$

$\cap, \cup>$, en donde P(A) es el conjunto potencia de cualquier conjunto A, es un retículo booleano. El álgebra de Lindebaum-Tarski del cálculo proposicional clásico es un retículo booleano. Una introducción elemental a la teoría de retículos distributivos en español es Hermes, 1963.

BIBLIOGRAFÍA

Beltrametti, E. y van Fraassen, B. C. (comps.) (1979), *Current Issues in Quantum Logic,* Plenum Press, New York-London.

Birkhoff, G. y von Neumann, J. (1936), «The Logic of Quantum Mechanics»: *Ann. Math., 37,* 823-843. Reproducido en C. A. Hooker (comp.), 1975.

Dalla Chiara, M. L. (1977), «Quantum Logic and physical modalities»: *Journal of Philosophical Logic, 6,* 391-404.

Dalla, Chiara, M. L. (1986), «Quantum Logic», en D. Gabbay y F. Guenthner (comps.), *Handbook of Philosophical Logic,* vol. III, Reidel, Dordrecht-Boston.

Dishkant, H. (1977), «Imbedding of the quantum logic in the modal system of Brouwer»: *Journal of Symbolic Logic, 42,* 321-328.

Finkelstein (1969), «Matter Space and Logic», en C. A. Hooker (comp.), 1979, 123-139.

Goldblatt, R. H. (1974), «Semantic analysis of orthologic»: *Journal of Phil. Logic, 3,* 19-35.

Goldblatt, R. H. (1984), «Orthomodularity is not elementary»: *Journal of Symbolic Logic, 49,* 401-404.

Hooker, C. A. (comp.) (1975-1979), *The Logico-Algebraic Approach to Quantum Mechanics,* vols. I, II, Reidel, Dordrecht.

Hermes, H. (1963), *La teoría de retículos y su aplicación a la lógica matemática,* Madrid.

Hardegree, G. (1977), «Reichenbach and the Logic of Quantum Mechanics»: *Synthese, 35,* 3-40.

Hardegree, G. (1979), «Charting the Labyrinth of Quantum Logics», en Beltrametti and van Fraassen (comps.), 1979.

Jauch, J. M. (1968), *Foundations of Quantum Mechanics,* Addison Wesley, Reading, Mass., Menlo Park, Ca.

Kochen, S. y Specker, E. P. (1967), «The Problem of Hidden Variables in Quantum Mechanics»: *Journal of Mathematics and Mechanics, 17,* 59-87. Reproducido en C. A. Hooker (comp.), 1975.

Piron, C. (1964), «Axiomatique quantique»: *Helvetica Physica Acta, 37,* 439-468.

Putnam, H. (1969), «Is logic empirical?», en *Boston Studies in the Philosophy of Science,* V, Reidel, Dordrecht. Reproducido en C. A. Hooker (comp.), 1975.

Rasiowa, H. (1974), *Algebraic Approach to Non-Classical Logics,* North-Holland, Amsterdam.

Reichenbach, H. (1944), *Philosophical Foundations of Quantum Mechanics,* Univ. of California Press, Los Angeles.

Stachow (1979), «Sequential quantum logic», en Beltrametti y van Fraassen (comps.), 1979, 173-192.

Strauss, M. (1937), «Mathematics as Logical Syntax - a Method to Formalize the Language of a Physical Theory»: *Erkenntnis, 7.* Reproducido en C. A. Hooker (comp.), 1975.

Van Fraassen, B. C. (1970), «On the extensión of Beth Semantics of Physical Theories»: *Philosophy of Science, 37,* 325-339.

Van Fraassen, B. C. (1974), «The Labyrinth of Quantum Logics», en R. Cohen y M. Wartofsky (comps.), *Boston Studies in the Philosophy of Science,* vol. 13, Reidel, Dordrecht.

LA LÓGICA DE LA RELEVANCIA

José M. Méndez

I. INTRODUCCIÓN

El origen de la lógica de la relevancia actual es, sin duda, el excelente artículo de W. Ackermann «Begründung einer strengen Implikation»[1] (Hay traducción española: cf. la bibliografía que cierra este trabajo; respecto de los antecedentes históricos, cf. Sylvan, 1988), pero su sistematización y desarrollo se deben a A. R. Anderson, N. D. Belnap y sus colaboradores. En «The pure calculus of entailment» (Anderson y Belnap, 1962), Anderson y Belnap establecen las bases filosóficas que validarían el intento de construir una lógica de la relevancia. En *Entailment* (vol. I, Anderson y Belnap, 1975; vol. II, Anderson, Belnap y Dunn, 1992) se incluyen todos los resultados sobre el tema hasta 1989 aproximadamente.

La motivación fundamental tras la lógica de la relevancia fue inicialmente filosófica: Anderson y Belnap querían definir una alternativa a la Lógica clásica en la formalización del discurso ordinario. El objetivo, como se ve, no puede calificarse de modesto. Y, así, la Lógica de la relevancia es, en la actualidad, un campo en continua expansión que, desbordando las (necesariamente desbordables) coordenadas iniciales, ha dado, y está dando lugar, a importantes desarrollos técnicos en álgebra intensional, teoría de sistemas formales, semántica de los mundos posibles no-estándar, teoría de la computación, teoría de conjuntos no clásica y filosofía de la lógica.

El objetivo del presente artículo es introducir al lector a todas estas cuestiones. Debido a limitaciones de espacio sobre las que no es preciso insistir, y, por otra parte, a la amplitud del tema, hemos centrado nuestra exposición en los puntos siguientes. En § II - § IV hemos intentado explicar la motivación subyacente a la Lógica de la relevancia. Comenzamos planteando el problema de las llamadas «paradojas del condicio-

nal clásico» que es el origen de la Lógica de la relevancia (§ II). Se examina a continuación la solución al problema propuesta por C. I. Lewis, el creador de la moderna lógica modal, que es, en nuestra opinión, el auténtico pionero en las investigaciones sobre la relevancia (§ III). Por último, estudiamos el origen, fundamentación y constitución del sistema **R** que es la solución al problema de las paradojas defendida por Anderson y Belnap.

En § V - § VII se expone con detalle este sistema, **R**, sin duda el más importante sistema de la lógica de la relevancia. Estos tres apartados, de carácter técnico, son, nos parece, un complemento imprescindible a los tres anteriores. Nuestro propósito al redactarlos, ha sido que el lector, además de conocer los fundamentos filosóficos de la Lógica de la relevancia, la *maneje* al menos en su parte proposicional o enunciativa, y tanto sintáctica (§ V) como semánticamente (§ VI, § VII). Por esta razón, sólo presuponemos un cierto conocimiento sintáctico de la lógica proposicional clásica y de la semántica estándar de la lógica modal (cfr. los artículos «Lógica clásica de primer orden» y «Lógica modal» en este mismo volumen). Por esta razón, también, hemos explicitado con todo detalle todas las pruebas salvo aquellas que proponemos al lector como ejercicio una vez que le hemos proporcionado previamente las claves para desarrollarlas.

En § VIII - § IX nos ocupamos de otros sistemas alternativos a **R** siempre en la línea iniciada por Anderson y Belnap. El tratamiento de todos ellos es como el empleado para el sistema **R**, pero, aun siguiendo las directrices de § V - § VII, hemos intentado ser breves. Entendemos que estos dos apartados son esenciales para un conocimiento cabal de la lógica de la relevancia en el sentido de Anderson y Belnap, pero, de todos modos, han sido redactados para que pueda prescindirse de ellos al leer este artículo. En tal caso, el presente trabajo sería una exposición del origen, fundamentos y estructura del sistema **R**.

Finalizamos, en § X, mencionando algunos resultados capitales sobre el tema no abordados en este artículo debido a su complejidad y, también, con algunas conclusiones sobre lo expuesto.

II. EL PROBLEMA: LAS PARADOJAS DEL CONDICIONAL MATERIAL

1. *Definición de lógica. Definición de lógica clásica*

La lógica se define tradicionalmente como la ciencia que se ocupa de *formalizar y sistematizar el concepto de inferencia (argumentación) deductiva correcta.* La tarea de la lógica consiste, por tanto, en definir el conjunto de todas las inferencias deductivas *correctas* que se corresponden con un determinado concepto de validez. Así, «inferencia correcta» e «inferencia válida» pueden considerarse, si el objetivo de la lógica se culmina, conceptos sinónimos.

Si nos circunscribimos al contexto de la lógica proposicional o enunciativa, la Lógica clásica puede caracterizarse por (i) bivalencia y (ii) funcionalidad de verdad. Es decir, la Lógica clásica se ocupa de definir el conjunto de inferencias cuya validez es *función de* (depende exclusivamente de) la *verdad* (entendida ésta en sentido bivalente) de los enunciados que las componen.

2. *Las paradojas del condicional material clásico*

La definición clásica de validez determina la interpretación de las conectivas proposicionales. Y, así, enunciados del tipo condicional $A \rightarrow B$ (Si *A*, entonces *B*) son falsos syss (si y sólo si) *A* (antecedente) es verdadero y *B* (consecuente) es falso. O, dicho contraposicionalmente, $A \rightarrow B$ es verdadero syss *A* es falso o bien *B* es verdadero. De esta interpretación del condicional se sigue que esquemas inferenciales tales como

1) $A \rightarrow (B \rightarrow A)$
o 2) $A \rightarrow (\neg A \rightarrow B)$

son válidos. Pero 1) y 2), interpretados extralógicamente, dan lugar, p. ej., a inferencias como

3) Si la Luna es un queso de bola, entonces $2+2=4$
y 4) Si $2+2 \neq 4$, entonces la Luna es un queso de bola.

(Sustitúyase «*A*» por «$2+2=4$», y «B» por «La Luna es un queso de bola». Entonces, 3) y 4) se siguen de 1) y 2) respectivamente, dado que «$2+2=4$» es una afirmación incuestionable).

Pues bien, 1) y 2) son dos ejemplos paradigmáticos de las denominadas, en sentido etimológico, «paradojas del condicional material (clásico)»: 1) y 2) apoyan la validez de, respectivamente, 3) y 4) que entran en conflicto, obviamente, con la idea intuitiva (ordinaria) de argumentación o inferencia válida. Por esta razón, el condicional clásico y, en consecuencia, la propia lógica clásica han sido tachados de paradójicos.

III. LA LÓGICA DE LA IMPLICACIÓN ESTRICTA DE C. I. LEWIS

1. *La lógica de la implicación de C. I. Lewis*

En la actualidad, C. I. Lewis es considerado con toda justicia el creador de la moderna *lógica modal;* es decir, la lógica de la *necesidad* y la *posibilidad* lógicas (cf. el capítulo «Lógica modal» de este volumen). Sin embargo, el propósito de Lewis fue, inicialmente, definir una lógica libre de paradojas como las comentadas más arriba; una lógica cuyo condicional se ajustara de modo más *estricto* al uso que hacemos en el lenguaje ordinario de la locución «Si ..., entonces ...».

Partiendo de una crítica a la lógica clásica semejante a la esbozada en el apartado anterior, Lewis concluye que las meras relaciones veritativo-

bivalentes entre el antecedente y el consecuente de un condicional son insuficientes para dar razón de su validez. Propone, por consiguiente, que un condicional se considere válido syss (si y sólo si) hay *relación de necesidad* entre antecedente y consecuente: syss es *imposible* (lógicamente) que el antecedente sea verdadero y el consecuente falso.

Siguiendo estas directrices, Lewis define en *Symbolic logic* (Lewis y Langford, 1932) cinco sistemas lógicos de la implicación estricta diferentes (desde entonces denominados **S1-S5**) que son otras tantas formulaciones del concepto «relación de necesidad entre antecedente y consecuente» de un condicional. Ha podido comprobarse posteriormente que las cinco alternativas propuestas por Lewis delimitan, de hecho, el espectro en el que en la actualidad cabe formalizar las nociones (esencialmente relativas) de necesidad y posibilidad lógicas más interesantes. Ahora bien, ¿cuáles fueron los resultados de Lewis respecto de la eliminación de las paradojas del condicional?, ¿alguno de sus sistemas (**S1-S5**) es la representación formal de la noción ordinaria reflejada en la locución «Si ..., entonces ...»?

2. *Las paradojas de la implicación estricta*

En ninguno de los sistemas de Lewis son válidos esquemas inferenciales como 1) y 2) (cf. § II.2), pero sí lo son, por ejemplo, en un sistema tan débil como **S2**:

5) $LA \prec (B \prec LA)$

o 6) $\neg MA \prec (A \prec B)$

De acuerdo con Lewis, 1) y 2) pueden interpretarse como sigue:

1') Si *A* es un enunciado verdadero, *A* es implicado por cualquier otro enunciado *B*.
2') Si *A* es un enunciado falso, *A* implica a cualquier otro enunciado *B*.

Así pues, 5) y 6) pueden interpretarse de modo paralelo:

5') Si *A* es un enunciado *necesariamente verdadero (LA), A* es *implicado estrictamente* ($\prec$) por cualquier otro enunciado *B*.
6') Si *A* es un enunciado *necesariamente falso* (imposible, $\neg MA$) *A implica estrictamente* ($\prec$) a cualquier otro enunciado *B*.

Por tanto, a pesar de que ninguno de los sistemas de Lewis daría lugar a consecuencias como las ejemplificadas en 3) y 4), sí darían lugar, p. ej., a

7) Si la Luna es un queso de bola, entonces $2+2=4$ ó $2+2\neq 4$.

y 8) Si $2+2=4$ y $2+2\neq 4$, entonces la Luna es un queso de bola.

Pues bien, 5) y 6) son dos ejemplos paradigmáticos de las denominadas «paradojas de la implicación estricta» porque 5) y 6), al igual que 1) y 2), originan condicionales extralógicos en los que el consecuente no tiene nada que ver con el antecedente, no se *sigue,* no *está implicado* por él. Es, entonces, obvio que la crítica de Lewis a la Lógica clásica es aplicable a sus propios sistemas: ya que todos contienen paradojas de la implicación, ninguno de ellos puede ser la representación formal del concepto de inferencia del lenguaje ordinario.

La conclusión del propio Lewis fue, sin embargo, distinta. Tras subrayar la diferencia entre las paradojas del condicional material y las de la implicación estricta (claramente ejemplificadas en 1') - 2') y 5') - 6'), respectivamente), afirmó que estas últimas eran ineliminables. En su opinión, eliminarlas equivaldría a eliminar también algunos principios esenciales (y no paradójicos) sobre el concepto de implicación, de modo que el sistema lógico resultante sería tan débil que, a la postre, carecería de interés por ser pragmáticamente inaplicable. Comprobamos en lo que sigue hasta qué punto esta opinión puede sostenerse.

IV. LA LÓGICA DE LA RELEVANCIA R

1. *La caracterización semántica de la relevancia*

1.1. Definición de «paradoja de la relevancia»

A fin de definir una lógica no paradójica (una lógica de la implicación estricta), Lewis tomó como punto de referencia expresiones paradójicas como 1) y 2), pero no dispuso de una definición precisa del término «paradoja de la implicación». Anderson y Belnap, en cambio, partieron de una definición precisa de dicho término que, en una primera aproximación, podría rezar como sigue:

> C1) Una expresión de forma condicional (una implicación) $A \rightarrow B$ es una *paradoja de la relevancia* si el contenido semántico de A y B es, eventualmente, disjunto.

De acuerdo con C1), 3), 4), 7) y 8) son, claramente, paradojas de la relevancia pues se establece relación de implicación entre proposiciones cuyo contenido semántico no está relacionado de modo alguno. Es decir, se establece relación de implicación entre antecedente y consecuente sin que aquel sea *relevante (pertinente)* para establecer éste: de la verdad o falsedad de la proposición «La Luna es un queso de bola» nada puede seguirse sobre la verdad o falsedad de la proposición «2 + 2 = 4». Así pues, la Lógica clásica y la Lógica de la implicación estricta son, para Anderson y Belnap, igualmente paradójicas sin que sea en absoluto esencial si ha de ser una verdad factual o bien una verdad necesaria el paradigma de las proposiciones implicadas por cualquier proposición. El hecho real-

mente esencial es que de una y otra lógica se sigue la posibilidad de establecer relaciones de implicación entre proposiciones que, en realidad, no mantienen tal relación.

1.2. La caracterización semántica de «relevancia»

Con C1) como criterio directriz, Anderson y Belnap inician la definición de una lógica proposicional no paradójica. El primer paso ha de ser, entonces, traducir formalmente C1). Pues bien, puesto que en lógica proposicional el contenido no lógico está representado por las variables proposicionales, la siguiente es la traducción formal de C1):

C2) Una fórmula de la forma $A \rightarrow B$ es una *paradoja de la relevancia,* si A y B no tienen en común al menos una variable proposicional.

De acuerdo con C2), 1), 2), 7) y 8) son, p. ej., paradojas de la relevancia, pues es fácil derivar a partir de ellas fórmulas que incumplen el requisito impuesto en C2) (p. ej., $B \rightarrow (A \rightarrow A)$, $\neg (A \rightarrow A) \rightarrow B$, $B \rightarrow L(A \rightarrow A)$, $\neg M(A \rightarrow A) \rightarrow B$ son derivables de 1), 2), 7) y 8) respectivamente, sustituyendo A por $A \rightarrow A$.

La definición C2) sugiere la siguiente caracterización semántica de relevancia:

Σ) L es una *lógica de la relevancia* si en todo teorema de L de la forma $A \rightarrow B$ A y B tienen al menos una variable proposicional en común.

El criterio Σ) es, por tanto, una condición necesaria de relevancia: si un sistema de lógica cuenta con la propiedad Σ), tenemos la garantía de que el antecedente y el consecuente de todo condicional derivable compartirán contenido semántico. Ahora bien, ¿cuál es la condición de suficiencia?, ¿cómo construir un sistema tal? Nos ocupamos de estas cuestiones en los apartados siguientes.

2. *La caracterización sintáctica de «relevancia»*

2.1. La definición de premisa «relevante»

Volvamos a fórmulas como 1). El problema de 1) y, en general, de las paradojas de la relevancia, es, desde un punto de vista semántico, que el antecedente nada tiene que ver con el consecuente. Desde el punto de vista sintáctico, esta deficiencia puede reformularse diciendo que el antecedente es *inútil* para probar el consecuente: operando, razonando únicamente con el antecedente nada podemos establecer sobre la derivabilidad del consecuente. Pues bien, precisamente es éste el punto de partida de Anderson y Belnap. A diferencia de Lewis, no exigen, para evitar las paradojas, que entre antecedente y consecuente se dé relación de necesi-

dad; exigen que el antecedente se *use* para probar el consecuente. No, por supuesto, que el antecedente *haya de usarse,* sea *imprescindible* para probar el consecuente (pues un mismo consecuente puede seguirse de diversos antecedentes), pero sí que pueda usarse, sí que para afirmar $A \rightarrow B$, A haya sido *usado* para demostrar B.

2.2. La caracterización semántica de «relevancia»

Esta noción intuitiva de «uso» puede traducirse, formalmente, en los siguientes términos:

C3) A se *usa* para probar B en una deducción dada Δ syss A es una premisa en una aplicación de una regla de inferencia en Δ.

Tomando como criterio C3), es obvio que, p. ej., 1), 2), 7) y 8) no pueden ser teoremas de ninguna lógica de la relevancia L pues, p. ej., de 1) se sigue fácilmente $B \rightarrow (A \rightarrow A)$, y, en la demostración de esta fórmula es claro que el antecedente no juega ningún papel en la demostración del consecuente. La definición C3) sugiere la siguiente caracterización sintáctica de relevancia:

S) L es una *lógica de la relevancia* si en todo teorema de L de la forma $A \rightarrow B$, A se usa para probar B.

3. *La lógica de la relevancia* ***R***

3.1. Insuficiencia de la caracterización sintáctica y de la caracterización semántica

Tenemos ya a nuestra disposición caracterizaciones sintáctica y semántica de «relevancia», ¿cómo definir a partir de ellas el sistema de la lógica de la relevancia? Supongamos que tomamos la caracterización semántica Σ) como directriz. El problema que se plantea es el siguiente: el conjunto de todos los condicionales no paradójicos en el sentido de C2) no es axiomatizable en ninguna lógica L. Comprobémoslo con un sencillo ejemplo. Sea L una lógica con los axiomas siguientes:

9) $(A \wedge B) \rightarrow A$
10) $A \rightarrow (B \rightarrow (A \wedge B))$
11) $(B \rightarrow C) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$

y con *modus ponens* (Si $\vdash A$ y $\vdash A \rightarrow B$, entonces $\vdash B$) como regla de derivación. Es obvio que 9), 10) y 11) son condicionales no paradójicos en el sentido de C2), y, sin embargo, es fácil probar que de este sistema se sigue el teorema

1) $A \rightarrow (B \rightarrow A)$

cuyo carácter paradójico ha sido pormenorizadamente discutido más arriba. La caracterización semántica Σ) es, por tanto, insuficiente para definir la lógica de la relevancia **R**.

Tomemos, pues, como directriz la caracterización sintáctica S). Desgraciadamente, nos topamos con un problema semejante: 9), 10) y 11) son no paradójicos en el sentido de C3) y, sin embargo, dan lugar a 1) que, evidentemente, sí lo es. La caracterización sintáctica S) es, en consecuencia, también insuficiente. ¿Es, entonces, imposible como Lewis sugería la construcción de la Lógica de la relevancia?

3.2. La lógica de la relevancia **R**

Hemos comprobado en el apartado anterior la insuficiencia de las caracterizaciones intuitivas (sintáctica y semántica) de «relevancia» propuestas por Anderson y Belnap. Es por esta razón por la que hemos hablado de caracterizaciones, y no de definiciones. Parece, por tanto, inevitable deducir la imposibilidad de definir una lógica de la relevancia a no ser que se propongan caracterizaciones sintáctica y semántica alternativas a las defendidas por Anderson y Belnap. Y, sin embargo, no es así. Dada la situación descrita, la estrategia de estos autores consiste, en esencia, en *corregir la insuficiencia del requisito del uso de la premisa con la que la exigencia de que las fórmulas derivables sean no paradójicas en el sentido semántico.* Con mayor precisión, la solución de Anderson y Belnap podría reformularse como sigue. Sean $R_1,\ldots,R_n$ el conjunto de reglas de derivación basadas en el criterio S). Pues bien, el sistema de la lógica de la relevancia **R** es el sistema lógico con la propiedad Σ) que equivale al máximo subconjunto posible de ese conjunto de reglas que no permitan la derivación de fórmulas paradójicas.

Anderson y Belnap demuestran que lo que hoy conocemos como *sistema de la lógica de la relevancia* **R** es un sistema equivalente a uno de los máximos subconjuntos posibles antes mencionados. Muestran también que el sistema **R** es, en contra de los augurios de Lewis, un sistema muy potente y que, en particular, no *está incluido* en ninguno de los sistemas definidos por el propio Lewis. Volveremos sobre estas afirmaciones al final de este trabajo. Por el momento, finalizamos este apartado describiendo con más detalle la solución propuesta por Anderson y Belnap.

La estrategia consiste en generalizar la caracterización de «teorema de la lógica de la relevancia» implícita en S) a

C4) Si $A_1\rightarrow(\ldots\rightarrow(A_n\rightarrow B)\ldots)$ es teorema, entonces *cada* $A_i (1 \leq i \leq n)$ se *usa* para probar B.

Demostraron entonces que el siguiente sistema implicativo **R**→ (un sistema es *implicativo* syss su única conectiva es →):

Axiomas: A1. $A\rightarrow A$

A2. $A\rightarrow((A\rightarrow B)\rightarrow B)$

A3. $(A\rightarrow(A\rightarrow B))\rightarrow(A\rightarrow B)$

A4. $(A\rightarrow B)\rightarrow((B\rightarrow C)\rightarrow(A\rightarrow C))$

Reglas de derivación: Modus ponens (Si $\vdash A$ y $\vdash A\to B$, entonces $\vdash B$), es *deductivamente completo* respecto de C4). Es decir, $A_1\to(\ldots\to(A_n\to B)\ldots)$ es teorema de **R→** syss cada $A_i (1\leq i\leq n)$ se usa en la prueba de *B*. Mostraron, además, que **R→** es un sistema relevante desde el punto de vista semántico. Concluyeron, por tanto, que **R→** es *el fragmento implicativo* de la Lógica de la relevancia.

Posteriormente, demostraron que los axiomas siguientes (para la conjunción (∧), disyunción (∨) y negación (¬)):

A5. $(A\wedge B)\to A\,/\,(A\wedge B)\to B$
A6. $((A\to B)\wedge(A\to C))\to(A\to(B\wedge C))$
A7. $A\to(A\vee B)\,/\,B\to(A\vee B)$
A8. $((A\to C)\wedge(B\to C)\to((A\vee B)\to C)$
A9. $(A\wedge(B\vee C))\to((A\wedge B)\vee C)$
A10. $A\to\neg\neg A$
A11. $(A\to\neg A)\to\neg A$
A12. $(\neg A\to B)\to(\neg B\to A)$

y la regla *Adjunción* (si $\vdash A$ y $\vdash B$, entonces $\vdash A\wedge B$) son compatibles con **R→** en el sentido de que se preserva el concepto de relevancia en sus dos sentidos, sintáctico y semántico. Anderson y Belnap concluyeron que el sistema anterior (A1-A12, *modus ponens* y *adjunción*), al que denominaron **R**, es la *Lógica de la relevancia*.

V. EL SISTEMA R DE LA LÓGICA DE LA RELEVANCIA

El lenguaje proposicional está compuesto por un conjunto enumerable de variables $P_o,\ldots, P_n\ldots$, las conectivas →, ∧, ∨, ¬, ↔ y los signos auxiliares (,). El conjunto de las fbf es el habitual. El bicondicional (↔) se define también al modo habitual.

1. *El sistema R*

Axiomas:

A1. $A\to A$
A2. $A\to((A\to B)\to B)$
A3. $(A\to(A\to B))\to(A\to B)$
A4. $(A\to B)\to((B\to C)\to(A\to C))$
A5. $(A\wedge B)\to A\,/\,(A\wedge B)\to B$
A6. $((A\to B)\wedge(A\to C))\to(A\to(B\wedge C))$
A7. $A\to(A\vee B)\,/\,A\to(B\vee A)$
A8. $((A\to C)\wedge(B\to C))\to((A\vee B)\to C)$
A9. $(A\wedge(B\vee C))\to((A\wedge B)\vee C)$
A10. $A\to\neg\neg A$
A11. $(A\to\neg A)\to\neg A$
A12. $(\neg A\to B)\to(\neg B\to A)$

Reglas de derivación:

Modus ponens: Si $\vdash A$ y $\vdash A\to B$, entonces $\vdash B$

y Adjunción: Si $\vdash A$ y $\vdash B$, entonces, $\vdash A\wedge B$.

A fin de que el lector pueda relacionar el sistema anterior con otros más conocidos, señalamos

Nota 1. Si en **R** sustituimos A1 y A2 por $A\to(B\to A)$, y eliminamos A12 el resultado es una axiomatización de la lógica proposicional intuicionista. (A9 no sería independiente.)

Nota 2. Si en **R** sustituimos A1 y A2 por $B\to(A\to A)$ el resultado es una axiomatización del sistema **S4** de Lewis con $\to$ (implicación estricta) como conectiva primitiva.

Nota 3. Si en **R** sustituimos A1 y A2 por $A\to(B\to A)$, el resultado es una axiomatización de la lógica proposicional clásica definida con $\to$, $\wedge$, $\vee$, $\neg$ como conectivas primitivas.

2. *Algunos teoremas característicos de* R

Anotamos una serie de teoremas característicos del sistema **R** con un esquema de prueba de cada uno de ellos. Al final de la lista desarrollamos con detalle alguna de estas pruebas.

T1. $((A\to A)\to B)\to B$ A1, A2
T2. $(A\to(B\to C))\to((A\to B)\to(A\to C))$ . . A2, A3, A4
T3. $(A\to(B\to C))\to((B\to(A\to C))$ A2, A4
T4. $(A\to B)\to((A\to(B\to C))\to(A\to C))$ . . T2, T3
T5. $(B\to C)\to((A\to B)\to(A\to C))$ A4, T3
T6. $(A\vee B)\leftrightarrow(B\vee A)$ A7, A8
T7. $(A\vee(B\vee C))\leftrightarrow((A\vee B)\vee C))$ A7, A8
T8. $A\leftrightarrow(A\vee A)$ A7; A1, A8
T9. $(A\wedge B)\leftrightarrow(B\wedge A)$ A5, A6
T10. $(A\wedge(B\wedge C))\leftrightarrow((A\wedge B)\wedge C)$ A5, A6
T11. $A\leftrightarrow(A\wedge A)$ A1, A6; A5
T12. $A\leftrightarrow(A\vee(A\wedge B))$ A7; A1, A5, A6
T13. $A\leftrightarrow(A\wedge(A\vee B))$ A1, A7, A8; A5
T14. $(A\vee(B\wedge C))\leftrightarrow((A\vee B)\wedge(A\vee C))$ A5-A8; A5-A9
T15. $(A\wedge(B\vee C))\leftrightarrow((A\wedge B)\vee(A\wedge C))$ A5-A9; A5-A8
T16. $(A\to(B\to C))\to((A\wedge B)\to C)$ A5, T4
T17. $(A\to(B\wedge C))\to((A\to B)\wedge(A\to C))$ . . . A5, A6
T18. $((A\to B)\vee(A\to C))\to(A\to(B\vee C))$ . . . A7, A8
T19. $((A\to C)\vee(B\to C))\to((A\wedge B)\to C)$ . . . A5, A8
T20. $((A\vee B)\to C)\to((A\to C)\wedge(B\to C))$ . . . A6, A7
T21. $\neg\neg A\to A$ A1, A12
T22. $(\neg A\to\neg B)\to(B\to A)$ A10, A12
T23. $(A\to B)\to(\neg B\to\neg A)$ A10, A12
T24. $(A\to\neg B)\to(B\to\neg A)$ A10, T23
T25. $(\neg A\to A)\to A$ A10, A11, T21

T26. $(A \to B) \to ((A \to \neg B) \to \neg A)$ A11, T23
T27. $(\neg A \to \neg B) \to ((\neg A \to B) \to A)$ . . T22, T25
T28. $(A \to B) \to ((\neg A \to B) \to B)$ T23, T25
T29. $\neg (A \vee B) \leftrightarrow (\neg A \wedge \neg B)$ A6, A7, T23; A5, A6, T24
T30. $\neg (A \wedge B) \leftrightarrow (\neg A \vee \neg B)$ A6, A7, A12; A5, A8, T23
T31. $(A \vee B) \leftrightarrow \neg (\neg A \wedge \neg B)$ A12, T24, T30
T32. $(A \wedge B) \leftrightarrow \neg (\neg A \vee \neg B)$ A12, T24, T30
T33. $(\neg A \to B) \to (A \vee B)$ A7, T23, T27
T34. $(A \to B) \to \neg (A \wedge \neg B)$ A5, T26
T35. $(A \to B) \to (\neg A \vee B)$ T21,T23
T36. $(A \wedge B) \to \neg (A \to \neg B)$ A6, A10, T24, T34
T37. $A \vee \neg A$ A7, A12, T23, T27
T38. $\neg (A \wedge \neg A)$ A5, T26

La mayoría de las pruebas son como las que daríamos en un sistema axiomático de lógica clásica formulado de modo igual o semejante al descrito en la Nota 3 (cfr. supra). Comprobémoslo con un par de ejemplos.

Observación 1. «1», «2», etc., designan en las pruebas que siguen los teoremas que aparecen en las líneas correspondientes. Las mismas consideraciones se aplican al uso de T1, T2, etc.

Observación 2. El uso de los teoremas implicativos T1 - T5 (y, en especial el de A4 - A5 (—transitividad del condicional—) en las pruebas subsiguientes no aparece reflejado, y así sucede en general, en la justificación esquemática que aparece a la derecha de cada teorema de la lista.

Ejemplo 1. T27. $(\neg A \to \neg B) \to ((\neg A \to B) \to A)$

Prueba.
1. $(B \to A) \to ((\neg A \to B) \to (\neg A \to A))$ T5
2. $(\neg A \to \neg B) \to (B \to A)$ T22
3. $(2) \to ((1) \to (A \to \neg B) \to ((\neg A \to B) \to (\neg A \to A)))$ . . A4
4. $(\neg A \to \neg B) \to ((\neg A \to B) \to (\neg A \to A))$ . . . . 2 MP, 1, 2, 3
5. $(\neg A \to A) \to A$ T22
6. $(5) \to (((\neg A \to B) \to (\neg A \to A)) \to ((\neg A \to B) \to A))$ . . A4
7. $((\neg A \to B) \to (\neg A \to A)) \to ((\neg A \to B) \to A)$ . . . . MP, 5, 6
8. $(4) \to ((7) \to \text{T27})$ A4
9. T27 . . . . MP, 4, 7, 8

Ejemplo 2. T29. $\neg (A \vee B) \leftrightarrow (\neg A \wedge \neg B)$

Prueba. (De izquierda a derecha; en el otro sentido, la prueba es semejante).
1. $A \to (A \vee B)$ A7
2. $B \to (A \vee B)$ A7
3. $(1) \to (\neg (A \vee B) \to \neg A)$ T23
4. $(2) \to (\neg (A \vee B) \to \neg A)$ T23
5. $\neg (A \vee B) \to \neg A$ . . . . MP, 1, 3
6. $\neg (A \vee B) \to \neg B$ . . . . MP, 2, 4

7. (5)∧(6) .... ADJ., 5, 6
8. (7)→T29 .... A6
9. T29 .... MP, 7, 8

3. *El Metateorema de Intercambio de los Equivalentes. El Metateorema de Deducción*

Una formulación habitual del Metateorema de Intercambio de los Equivalentes (M.I.) es la siguiente:

Si $\vdash A \leftrightarrow B$, entonces $\vdash C(A) \leftrightarrow C(A/B)$

Su significado vendría a ser éste:
si dos fórmulas son sintácticamente equivalentes, entonces son intercambiables dentro de cualquier contexto (fbf) en el que aparezcan. En Lógica clásica, M.I. se prueba por inducción sobre la complejidad (grado lógico) de $C(A)$. Para mostrar su validez basta con las siguientes reglas:

a) Si $\vdash A \rightarrow B$, $\vdash (B \rightarrow C) \rightarrow (A \rightarrow C)$
b) Si $\vdash A \rightarrow B$, $\vdash (C \rightarrow A) \rightarrow (C \rightarrow B)$
c) Si $\vdash A \rightarrow B$, $\vdash \neg B \rightarrow \neg A$

Como a), b) y c) son, obviamente, reglas derivadas de **R**, M. I. se prueba para **R** exactamente igual que para la Lógica clásica.

Por otro lado, el Metateorema de Deducción clásico (M.D.) rezaría

(M.D.) Si $A_1, \ldots, A_n \vdash B$, entonces $A_1, \ldots, A_{n-1} \vdash A_n \rightarrow B$

En Lógica de la Relevancia en general, y en el sistema **R** en particular, se modifica como sigue (Metateorema de Deducción Relevante (M.D.r.)):

(M.D.r.) Si $A_1, \ldots, A_n \vdash B$, y cada $A_i (1 \leq i \leq n)$ se usa en la demostración de B, entonces $A_1, \ldots, A_{n-1} \vdash A_n \rightarrow B$

Hemos explicado con detalle más arriba (cf. §IV.2) las razones que justifican esta modificación. La prueba de M.D.r. es, en esencia, semejante a la prueba de M.D., pero se requieren algunas especificaciones técnicas en las que no podemos entrar aquí (cf. Dunn, 1986).

VI. SEMÁNTICA PARA R

Describimos a continuación la semántica para **R** definida por Routley y Meyer (cf. Routley y Meyer, 1973; Routley y otros, 1984). Semánticas esencialmente equivalentes son las propuestas por A. Urquhart y K. Fine (cf. Urquhart, 1972; Fine, 1974). La semántica de Routley y Meyer para R es semejante a la semántica kripkeana para la lógica modal (cf. el capítulo «Lógica Modal» de este mismo volumen).

Como se sabe, una «estructura kripkeana» es una estructura del tipo $<O, K, R>$ donde *K es un conjunto,* $O \in K$, y R es una relación binaria definida en K. Desde un punto de vista intuitivo, K es el conjunto de los mundos posibles, O es el «mundo actual» y R representa la relación de

accesibilidad entre mundos posibles. Para definir un «modelo kripkeano» añadimos a una estructura kripkeana una función de evaluación con la que se valora cada fórmula bien formada en cada mundo posible conforme a requisitos que respecto de las conectivas $\rightarrow$, $\wedge$, $\vee$, $\leftrightarrow$ y $\neg$ son exactamente igual a los clásicos.

Pues bien, las estructuras para **R** son estructuras del tipo $<O, K, R, *>$ donde *O, K* y *R* son como antes y * es una operación monaria definida en *K*. Las diferencias respecto de las estructuras kripkeanas estándar pueden resumirse en lo siguiente:

i) El conjunto *K* de los mundos posibles puede contener elementos inconsistentes o incompletos o inconsistentes e incompletos.
ii) *R* es una relación ternaria. Así, si en los modelos estándar «*Rab*» puede leerse «el mundo posible *a* tiene acceso al mundo posible *b*», en los modelos para **R**, «*Rabc*» podría leerse «el mundo posible *c* es accesible desde *a* y *b*». Naturalmente, el hecho de que *R* sea una relación ternaria, dificulta, a diferencia de lo que ocurre en los modelos estándar, la elección de las propiedades que cabe imponerle.
iii) La operación * proporciona una, diríamos, «imagen inversa» de cada mundo posible. Para cada mundo posible *a*, a^* contiene todas las fórmulas negativas cuyos argumentos están en *a;* o, dicho de otro modo, a^* carece de los argumentos de las fórmulas negativas que están en *a*.
iv) Las diferencias respecto de la semántica estándar expresadas en los dos puntos anteriores tienen como objetivo valorar el condicional y la negación de forma alternativa (cfr. *infra*) con vista a falsar las paradojas de la relevancia.

A continuación exponemos con detalle esta semántica, y probamos el Teorema de Consistencia.

1. *Estructuras modelo, modelos, validez*

Def. 1. Una *estructura modelo para* ***R*** *(EMR)* es una estructura del tipo $<O, K, R, *>$ donde

i) $K \neq \varphi$ (*K* es un conjunto no vacío)
ii) $O \subseteq K$ (*O* es un subconjunto de *K*)
iii) $R \subseteq K^3$ (*R* es una relación ternaria definida en *K*)
iv) $* \subseteq K$ (* es una operación monaria definida en *K*)

tales que las siguientes definiciones y postulados se cumplen para cualesquiera *a, b, c, d,* $\in K$ (los cuantificadores tienen como dominio *K;* «$\Rightarrow$» puede leerse «si ..., entonces ...»; es decir, como un condicional metalingüístico).

d1. $a \leq b =_{df} (x \in O$ y $Rxab)$
d2. $R^2abcd =_{df} \exists x(x \in O$ y $Rabx$ y $Rxcd)$

P1. $a \leq a$
P2. $a \leq b$ y $Rbcd \Rightarrow Racd$
P3. $Rabc \Rightarrow Rbac$
P4. $R^2abcd \Rightarrow \exists x(Racx$ y $Rbxd)$
P5. $Raaa$
P6. a^*
P7. $Rabc \Rightarrow Rab^*c^*$

Def. 2. Un *modelo para R (MR)* es una estructura del tipo <*O*, *K*, *R*, *,
ción de verdad con la que asignamos 1 ó 0 a cada fbf en cada mundo posible de acuerdo con las restricciones siguientes (para cualesquiera variable proposicional *p*, fbf *A*, *B* y $a \in K$):

i) $a \leq b$ y $v(A, a) = 1 \Rightarrow v(B, b) = 1$
ii) $V(A \wedge B) = 1$ syss $v(A, a) = 1$ y $v(B, a) = 1$
iii) $v(A \vee B) = 1$ syss $v(A, a) = 1$ ó $v(B, a) = 1$
iv) $v(A \rightarrow B) = 1$ syss para todo *b*, *c*, $\in K$, *Rabc* y $v(A, b) = 1 \Rightarrow v(B, c) = 1$
v) $v(\neg A, a) = 1$ syss $v(A, a^*) =$

Def. 3. *A* es *válida* ($\vDash$A) syss para todo modelo *M* y $a \in K$ tal que $a \in O$, $v(A, a) = 1$.

2. *Teorema de consistencia*

Para simplificar la prueba del Teorema de Consistencia, utilizamos la siguiente definición:

Def. 4. *A implica semánticamente a B en* **R** syss para todo *MR* y $a \in K$, $v(A, a) = 1 \Rightarrow v(B, a) = 1$
y los siguientes Lemas:

Lema 1. *A* implica semánticamente a *B* syss $\vDash A \rightarrow B$.

Prueba. Considérese cualquier *MR* en el que $a \in O$ y $b, c \in K$. Entonces,

i) Si *A* implica semánticamente a *B* en **R**, entonces $\vDash A \rightarrow$ B.
Suponemos, pues,
1. *A* implica semánticamente a *B* en **R** . . . . . Hip.
2. *Rabc* y $v(A, b) = 1$ . . . . Hip.
para demostrar, de acuerdo con DF2(iv) y DF3, $v(B, c) = 1$. Entonces,
3. $a \leq b$ . . . . 2, d1, pues $a \in O$
4. $v(A, b) = 1$ . . . . 2, 3, Def. 2(i)
5. $v(\mathrm{B}, \mathrm{b}) = 1$ . . . . 1, 4, Def. 4

ii) Si $\vDash A \rightarrow B$, entonces *A* implica semánticamente a *B* en **R**.
1. $\vDash A \rightarrow B$ . . . . Hip.
2. $v(A, b) = 1$ . . . . Hip.
De acuerdo con Def. 4, hemos de demostrar ahora $v(B, b) = 1$. Entonces,
3. *Rxbb* P1, d1
4. $v(A \rightarrow B, x) = 1$ . . . . 1, Def. 3, pues $x \in O$
5. $v(B, b) = 1$ . . . . 2, 3, 4, Def. 2(iv)

Lema 2. $v(A, a)=1$ syss $v(\neg A, a^*)=0$

Prueba. $v(A, a)=1$ syss $v(A, a^{**})=1$ (por P6) syss $v(\neg A, a^*)=0$ (por Def. 2 (v)).

Lema 3. Los postulados siguientes:

P5(a). $R^2abcd \Rightarrow R^2acbd$
P5(b). $Rabc \Rightarrow R^2abbc$
P5(c). Raa^*a

se cumplen en todas las *EMR*.

Prueba. (i) P5(a) $R^2abcd \Rightarrow R^2acbd$

1. R^2abcd .... Hip.
2. $Racx$ y $Rbxc$ 1, P4
3. $Racx$ y $Rxbd$ 2, P3
4. R^2acbd 3, d2

ii) P5(b) $Rabc \Rightarrow R^2abbc$

1. $Rabc$ .... Hip.
2. $Rbac$ 1, P3
3. $Rbbb$ P5
4. R^2bbac 2, 3, d2
5. R^2babc .... 4, P5(a)
6. $Rbax$ y $Rxbc$ 5, d2
7. $Rabx$ y $Rxbc$ 6, P3
8. R^2abbc 7, d2

iii) P5(c) Raa^*a

1. $Ra^*a^*a^*$ P5
2. $Ra^*a^{**}a^{**}$ 1, P8
3. Ra^*aa 2, P6
4. Raa^*a 3, P3

Podemos ahora probar el

Teorema de consistencia (semántica)

Si $\vdash A$, entonces $\models A$ (si A es teorema, entonces A es válida)

Prueba. La prueba consiste, como es habitual, en demostrar que los axiomas son válidos y que las reglas de derivación preservan la validez. Pues bien, como todos los axiomas son de la forma $A \rightarrow B$, por el Lema 1 basta probar que en cualquier *MR* y para cualquier $a \in K$, $v(A, a)=1 \Rightarrow v(B, a)=1$. Entonces, A1, A5, A6, A7, A8, A9, Modus ponens y Adjunción se siguen inmediatamente de Def. 2. Por otro lado, A2, A3, A4, A10, A11 y A12 se siguen fácilmente de, respectivamente, P3, P5(b), P4, Lema 2, P5(c) y P8. Desarrollamos a continuación algunas de estas pruebas. Procederemos por reducción al absurdo. Es decir, supondremos que el axioma del caso no es válido. Como todos los axiomas son de la forma $A \rightarrow B$, esto significa por el Lema 1 que para algún $a \in O$ en

algún MR, $v(A, a)=1$ y $v(B, a)=0$. Demostraremos que de tal suposición se sigue una contradicción:

A4. $(A\to B)\to((B\to C)\to(A\to C))$

1. (i) $v(A\to B, a)=1$
   (ii) $v((B\to C)\to(A\to C), a)=0$ ...... Hip.
2. (i) $Rabc$
   (ii) $v(B\to C, b)=1$
   (iii) $v(A\to C, c)=0$ .... 1(ii), Def. 2(iv)
3. (i) $Rcde$
   (ii) $v(A,d)=1$
   (iii) $v(C,e)=0$ .... 2(iii), Def. 2(iv)
4. R^2abde .... 2(i), 3(i), d2
5. (i) $Radx$
   (ii) $Rbxe$ 4, P4
6. $v(B, x)=1$ .... 1(i), 3(ii), 5(i), Def. 2(iv)
7. $v(C,e)=1$ .... 2(ii), 5(ii), 6

Pero 7 y 3(iii) se contradicen. Por tanto, A4 es válido.

A10. $A\to\neg\neg A$.

1. (i) $v(A, a)=1$
   (ii) $v(\neg\neg A)=0$ .... Hip.
2. $v(\neg A, a^*$
3. $v(\neg\neg A, a)=1$ .... 2, Def. 2) v)

Pero 3 y 1(ii) se contradicen. Por tanto, $A10$ es válido.

A12. $(\neg A\to B)\to(\neg B\to A)$

1. (i) $v(\neg A\to b, a)=1$
   (ii) $v(\neg B\to A)=0$ .... Hip.
2. (i) $Rabc$
   (ii) $v(B, b)=1$
   (iii) $v(A, c)=0$ .... 1(ii), Def. 2(iv)
3. Rac^*b
4. $v(\neg A, c^*)=1$ .... 2(iii), Lema 2
5. $v(B, b^*)=1$ .... 1(i), 3, 4, Def. 2(iv)
6. $v(\neg B, b)=0$ .... 5, Def. 2(v)

Pero 6 y 2(ii) se contradicen. Por tanto, A12 es válido.

Nota 1. La cláusula (ii) de la Def. 2 se restringe normalmente al caso de las variables proposicionales, y, entonces, rezaría:

Def. 2(ii)'. $a\leq b$ y $v(p, a)=1\Rightarrow v(p, b)=1$

El lector puede demostrar sin dificultad que Def. 2(ii)' se extiende a Def. 2(ii) por inducción sobre el grado lógico (la longitud, la complejidad) de las fbf.

Nota 2. Hemos utilizado «modelos no-reducidos» para **R** en lugar de los «modelos reducidos» (es decir, modelos sin un «mundo actual») (cfr. Routley y otros, 1982). La razón es que así podemos exponer de

modo unificado la semántica para otras lógicas de la relevancia alternativas.

3. *Algunos conocidos teoremas de la lógica proposicional clásica que no son teoremas de* **R**

1. $A \rightarrow (B \rightarrow A)$
2. $((A \rightarrow B) \rightarrow C) \rightarrow (B \rightarrow C)$
3. $((A \rightarrow B) \rightarrow B) \rightarrow ((B \rightarrow A) \rightarrow A)$
4. $((A \rightarrow B) \rightarrow A) \rightarrow A$
5. $(A \rightarrow C) \rightarrow ((B \rightarrow C) \rightarrow ((A \vee B) \rightarrow C)$
6. $(A \rightarrow B) \rightarrow ((A \rightarrow C) \rightarrow (A \rightarrow (B \wedge C))$
7. $A \rightarrow (B \rightarrow (A \wedge B))$
8. $((A \wedge B) \rightarrow C) \rightarrow (A \rightarrow (B \rightarrow C))$
9. $(A \rightarrow B) \rightarrow ((A \wedge C) \rightarrow (B \wedge C))$
10. $(A \rightarrow B) \rightarrow ((A \vee C) \rightarrow (B \vee C))$
11. $(A \rightarrow (B \vee C)) \rightarrow ((A \rightarrow B) \vee (A \rightarrow C))$
12. $((A \wedge B) \rightarrow C) \rightarrow ((A \rightarrow C) \vee (B \rightarrow C))$
13. $(A \rightarrow B) \vee (B \rightarrow A)$
14. $\neg A \rightarrow (A \rightarrow B)$
15. $\neg (A \rightarrow \neg B) \rightarrow (A \wedge B)$
16. $(A \vee B) \rightarrow (\neg A \rightarrow B)$
17. $(\neg A \vee B) \rightarrow (A \rightarrow B)$
18. $(A \wedge \neg B) \rightarrow (A \rightarrow B)$

A pesar de la indecidibilidad de **R** (cfr. §X) sugerimos al lector como ejercicio dos posibilidades de mostrar que las fórmulas anteriores no son teoremas de **R:**

(I) Sintaxis.

i) No es difícil demostrar que si añadimos a **R** cualquiera de (1)-(4), (7)-(8) ó (14)-(18), entonces $A \rightarrow (B \rightarrow A)$ es demostrable. Pero el resultado de añadir $A \rightarrow (B \rightarrow A)$ a **R** es, evidentemente, la Lógica clásica.

ii) **R** en unión de (5), (6), (9) ó (10) da lugar a teoremas como $(A \rightarrow B) \rightarrow (A \rightarrow A)$ ó $(B \rightarrow A) \rightarrow (A \rightarrow A)$; **R** en unión de (11), (12) ó (13) da lugar a $(A \rightarrow B) \vee (B \rightarrow A)$ como teorema. Como hemos discutido más arriba, uno y otro resultado son indeseables.

(II) Semántica.

Siguiendo la estrategia utilizada en la demostración de la validez de los axiomas de **R,** no es difícil, en la mayoría de los casos, construir modelos que muestran la no-validez de (1)-(18). Por ejemplo:

i) (1) $A \rightarrow (B \rightarrow A)$ se falsa en un MR en el que (a) $K = \{a, b, c\}$, (b) $O = K$, (c) $Rabc$, (d) $v(A, a) = 0$, $v(A, b) = 1$, $v(C, c) = 0$.

ii) (14) $\neg A \rightarrow (A \rightarrow B)$ se falsa en un MR en el que (a) $K = \{a, b, c\}$, (b) $O = K$, (c) $Rabc$, (d) $v(A, a^*$

iii) (15) $\neg (A \rightarrow \neg B) \rightarrow (A \wedge B)$ se falsa en un MR en el que (a) $K = \{a, b, c\}$ (b) $O = K$, (c) Ra^*bc, (d) $v(A, b) = 1$, $v(B, b^*) = 0$, $v(A, a) = 0$.

Nota. En cada uno de los modelos anteriores se especifica lo estrictamente necesario para falsar las fórmulas del caso. Lo no especificado es indiferente siempre que sea compatible con los requisitos que definen un *MR*. Así, p. ej., en i) es indiferente el valor de *A* en *b* ó si se dan o no entre *a*, *b* y *c* otras relaciones que *Rabc*.

VII. COMPLETUD DE R

Damos a continuación una prueba de completud de **R** respecto de la semántica descrita en § VI. La prueba es, en esencia, una prueba tipo Henkin. Es decir, definiremos el «modelo canónico» (cf. Def. 2), y probaremos que todas las fbf que no sean teoremas son falsas en algún mundo posible del modelo canónico. En la construcción del modelo canónico, utilizaremos el método de las teorías «primas»; es decir, interpretaremos canónicamente el conjunto *K* como el conjunto de todas las teorías primas (cf. Def. 1). Comenzamos con algunas definiciones.

1. *Definiciones. Modelo canónico*

Def. 1

i) Conjunto de fórmulas cerrado por la implicación ($\mapsto$).
a está cerrado por la implicación syss si $\vdash A \rightarrow B$ y $A \in a$, entonces $B \in a$.

ii) Conjunto de fórmulas cerrado por la adjunción (&).
a está cerrado por la adjunción syss si $A \in a$ y $B \in a$, entonces $A \wedge B \in a$.

iii) Teoría.
a es una *teoría* syss a está cerrada por $\mapsto$ y por &.

iv) Teoría normal.
a es una teoría normal syss (a) *a* es una teoría y (b) si $\vdash A$, entonces $A \in a$.

v) Teoría prima.
a es una *teoría prima* syss (a) *a* es una teoría y (b) si $A \vee B \in a$, entonces $A \in a$ ó $B \in a$.

Def. 2. Estructura modelo canónica para **R**.
Una *estructura modelo canónica para* **R** *(EMR)* es cualquier estructura del tipo $\langle O^c, K^c, R^c, {}^{*c} \rangle$ donde

i) K^c es el conjunto de todas las teorías primas.

ii) O^c es el conjunto de todas las teorías primas normales.

iii) R^c se define en K^c como sigue: para cualesquiera $a, b, c \in K^c$, $R^c abc$ syss si $A \rightarrow B \in a$ y $A \in b$, entonces $B \in c$.

iv) ^{*c} se define en K^c como sigue: para cualquier $a \in K^c$, $a^{*c} = \{A \mid \neg A \notin a\}$ (Es decir, $A \in a^{*c}$ syss $\neg A \notin a$).

2. *Lemas previos*

Probamos a continuación tres lemas esenciales para la demostración, en el apartado siguiente, de que el modelo canónico es, en efecto, un modelo.

Lema 1. Si $\nvdash A$, entonces hay una teoría T tal que $T \in K^c$ y $A \notin T$.

Prueba. En la demostración que sigue utilizamos el Lema de Zorn que reza: Todo conjunto no vacío en el que cada cadena (es decir, cada subconjunto totalmente ordenado) tenga un límite superior tiene un elemento máximo.

Hay pruebas alternativas (técnicamente más complicadas) utilizando el Lema de Lindenbaum (cfr. Routley y otros, 1982, cap. V).

Pasamos ya a la demostración del Lema 1. Considérese la lógica **R** como el conjunto de sus teoremas. **R** es una teoría (la mínima) normal sin A. Ordénense todas las teorías normales sin A con la inclusión conjuntista. Es evidente que el Lema de Zorn puede aplicarse, y, por tanto, hay una teoría normal y *máxima* T sin A ($A \notin T$) (que T es una teoría normal y máxima sin A significa: si T' es una teoría y $T \subset T'$ como T' es necesariamente normal, $A \in T'$).

Demostraremos a continuación que T es prima. Supongamos que no lo es. Entonces, para algunas fbf $B, C,$ tenemos $B \vee C \in T$, $B \notin T$, $C \in T$ (cfr. Def. 1 (iv)).

Definimos:

$$[T, B] = \{E \mid \exists D(D \in T, \ \vdash (B \wedge D) \rightarrow E)\}$$
$$[T, C] = \{E \mid \exists D(D \in T, \ \vdash (C \wedge D) \rightarrow E)\}$$

Es decir, $[T, B]$ es el conjunto de todas las fórmulas E tales que $(B \wedge D) \rightarrow E$ es teorema de **R**, siendo D una fórmula cualquiera de T; $[T, C]$ se puede interpretar de modo semejante. Pues bien, probamos ahora

i) $[T, B]$ y $[T, C]$ están cerrados por la implicación.

Prueba. Probamos que $[T, B]$ está cerrado por la implicación; la prueba de que $[T, C]$ también lo está es similar. Supongamos (cf. Def. 1 (i)), entonces,

1. $\vdash D \rightarrow E$ .... Hip.
2. $\vdash D \in [T, B]$ .... Hip.
3. $\vdash (B \wedge F) \rightarrow D (F \in T)$ .... 2, Def. de $[T, B]$
4. $\vdash (B \wedge F) \rightarrow E (F \in T)$ 1, 3, A4

y, como había que demostrar,

5. $E \in [T, B]$ .... 4, Def de $[T, B]$

ii) $[T, B]$ y $[T, C]$ están cerrados por la adjunción.

Prueba. Tomando como pauta la demostración de (i), el lector puede demostrar sin dificultad (úsense algunos teoremas elementales sobre la conjunción, cf. §v.2) que $[T, B]$ está cerrado por la adjunción; la prueba de que $[T, C]$ también lo está es semejante.

iii) $T \subset [T, B]$, $T \subset [T, C]$

Prueba. Demostraremos que T está incluida estrictamente en $[T, B]$; la prueba de que $[T, C]$ también lo está es semejante.

a) $T \subseteq [T, B]$ (si $D \in T$, entonces $D \in [T, B]$)
Prueba.
1. $\vdash (B \wedge D) \rightarrow D$ A5
2. $D \in [T, B]$ .... 1, Def. de $[T, B]$

b) $T \subset [T, B]$ (Hay, al menos, una fbf D tal que $D \in [T, B]$, pero $D \notin T$).
Prueba. La fbf aludida es, evidentemente, B. Por hipótesis, $B \notin T$, pero sea $E \in T$ (T es normal: no es vacía). Entonces,
1. $\vdash (B \wedge E) \rightarrow B$ A5
2. $B \in [T, B]$ .... 1, Def. de $[T, B]$

iv) $A \in [T, B]$, $A \in [T, C]$.
Prueba. T es la teoría normal máxima sin A. Como $[T, B]$ y $[T, C]$ son teorías normales que incluyen estrictamente a T((i), (ii), (iii)), entonces, $A \in [T, B]$ y $A \in [T, C]$.

v) $A \in T$.
Prueba. Utilizando algunos teoremas elementales sobre la conjunción, la disyunción y las relaciones entre ambas, el lector puede demostrar sin dificultad que si $A \in [T, B]$ y $A \in [T, C]$ (iv), entonces, $A \in T$.

La prueba del Lema 1 es ahora inmediata: Hemos construido una teoría normal T tal que $A \notin T$; hemos supuesto (hipótesis de reducción al absurdo) que T no es prima. De esta suposición se sigue, como hemos comprobado en (v), que $A \in T$. En consecuencia, T es una teoría normal y prima tal que $A \notin T$.

Lema 2. Para cualesquiera a, $b \in K^c$, $a \leq^c b$ syss $a \subseteq b$.
Prueba. Supóngase $a \leq^c b$. Por d1 y Def. 2 (iv), $R^c xab$ para algún $x \in O^c$. Por A1, $A \rightarrow A \in x$. Por tanto, si $A \in a$, entonces $A \in b$ (cf. Def. 2 (iv)). En el otro sentido, supóngase $a \subseteq b$. Como a es teoría, $R^c\mathbf{R}aa$ (a esta cerrada por $\mapsto$; **R** es el sistema **R**, el conjunto de los teoremas de **R**); como $a \subseteq b$, $R^c\mathbf{R}ab$, es decir, $\exists x(x$ es normal y $R^c xab)$ (**R** es normal). Queda por demostrar que x puede extenderse a una teoría prima z tal que $z \in O^c$ y $R^c zab$, pues así, $a \leq^c b$, como hay que demostrar (cf. d1).

Considérese, por tanto, el conjunto de todas las teorías normales Y tales que $x \subseteq y$ y $R^c yab$. Por el Lema de Zorn, hay un elemento máximo z en Y tal que $x \subseteq z$ y $R^c zab$. Demostraremos a continuación que z es prima. Supongamos que no lo es (reducción al absurdo). Entonces, $A \vee B \in z$, $A \notin z$, $B \notin z$ para algunas fbf A, B. Defínanse los conjuntos $[z, A]$ y $[z, B]$ de modo similar a como se definieron $[T, B]$ y $[T, C]$ en la prueba del Lema 1. Siguiendo la estrategia de prueba de este Lema, el lector no tendrá dificultad en demostrar que $[z, A]$ y $[z, B]$ son teorías normales que incluyen estrictamente a z. Como z es la teoría máxima tal que $x \subseteq z$ y $R^c zab$, se concluye (cf. la introducción al Lema 1)

1. No-$R^c[z, A]\ ab$
2. No-$R^c[z, B]\ ab$

y, entonces, tenemos, por 1 y Def. 2 (iv),

3. $C \rightarrow D \in [z, A]$
4. $C \in a$
5. $D \notin b$

por 2 y Def. 2 (iv),

6. $C' \rightarrow D'$
7. $C' \in a$
8. $D' \notin b$

por 3 y Def. de $[z, A]$,

9. $\vdash (A \wedge F) \rightarrow (C \rightarrow D)\ (F \in z)$

por 6 y Def. de $[z, B]$,

10. $\vdash (A \wedge F') \rightarrow (C' \rightarrow D')\ (F' \in z)$.

Ahora, dado que $((A \rightarrow C) \wedge (B \rightarrow D)) \rightarrow ((A \vee B) \rightarrow (C \vee D))$ es teorema de **R**, tenemos por 9 y 10,

11. $((A \wedge F) \vee (A \wedge F')) \rightarrow ((C \rightarrow D) \vee (C' \rightarrow D'))$

y, por las propiedades distributivas,

12. $((A \vee B) \wedge (F \wedge F')) \rightarrow ((C \rightarrow D) \vee (C' \rightarrow D'))$.

Pero como z está cerrada por & (z es teoría) y $F \in z$, $F' \in z$, entonces $F \wedge F' \in z$; por la misma razón, como $A \vee B \in z$,

13. $((A \vee B) \wedge (F \wedge F')) \in z$.

Como z está cerrada por $\mapsto$, por 12 y 13,

14. $((C \rightarrow D) \vee (C' \rightarrow D')) \in z$

Pero el siguiente es teorema de **R**

15. $\vdash ((C \rightarrow D) \vee (C' \rightarrow D')) \rightarrow ((C \wedge C') \rightarrow (D \vee D'))$

y, entonces, por 14 y 15 (z está cerrada por $\mapsto$),

16. $((C \wedge C') \rightarrow (D \vee D')) \in z$.

Por último,

17. $(D \vee D') \in z$

pues $(C \wedge C') \in C$ (a y $C' \in a$, y a está cerrada por & pues es teoría) y $R^c zab$ (cfr. Def. 2 (iv)).

Pero dado que b es prima ($b \in K^c$), de (16) se sigue $D \in b$ ó $D' \in b$ que contradice lo afirmado en las líneas 5 y 8 ($D \notin b$, $D' \notin b$). En consecuencia, z es una teoría prima. Como $x \subseteq z$ y x es normal, entonces $z \in O^c$ (z es normal) y $R^c zab$, *i.e*, $a \leq {}^c b$, con lo que finaliza la demostración (de la segunda parte) del Lema 2.

Lema 3. ^{*c} es una operación en K^c; es decir, si a es una teoría prima, entonces a^{*c} también lo es.

Prueba. Tenemos que demostrar que si a está cerrada por $\mapsto$, & y es prima, entonces a^{*c} también.

i) a^{*c} está cerrada por $\mapsto$.

Prueba. El lector lo demostrará sin dificultad siguiendo la estrategia de reducción al absurdo y utilizando uno de los teoremas de contraposición (T23).

ii) a^{*c} está cerrada por &.

Prueba.

1. $A \in a^{*c}$ .... Hip.
2. $B \in a^{*c}$ .... Hip.
3. $A \wedge B \notin a^{*c}$ .... Hip. (red. ad abs.)
4. $\neg (A \wedge B) \in a$ ..... 3, Def. 2(v)
5. $\neg A \vee \neg B \in a$ ... 4, T30 (a está cerrada por $\mapsto$, a es teoría)
6. $\neg A \in a$ ó $\neg B \in a$. 5, pues a es prima
7. $\neg A \notin a$ y $\neg B \notin a$. 1, 2, Def. 2(v)

pero (6) y (7) se contradicen. Por tanto, a^{*c} está cerrada por & cuando a es una teoría prima.

iii) a^{*c} es prima.

Prueba. Semejante a la de ii) utilizando ahora T29.

3. *Teorema de Completud*

Probamos a continuación que el modelo canónico es, en efecto, un modelo de donde se sigue ya con facilidad el Teorema de Completud.

Lema 4. La estructura modelo canónica es, en efecto, una estructura modelo.

Prueba. Tenemos que demostrar que la estructura establecida en Def. 2 cumple con los requisitos generales de toda estructura modelo para **R** definidos en §VI.1. Pues bien, por el Lema 1 se cumplen los requisitos (i), (ii) y (iii). Ahora, dado que R^c es, obviamente una relación ternaria definida en K^c (requisito (iv)), y $*^c$, por el Lema 3, una operación monaria (requisito (vi)), sólo resta por demostrar que los postulados P1-P7 se cumplen cuando se entienden canónicamente de acuerdo con Def. 2 (iv), (v).

P1. $a \leq^c a$ .... Lema 2

P2. si $a \leq^c b$ y $R^c bcd$, entonces $R^c acd$.

Prueba.

1. $a \leq^c b$ .... Hip.
2. $R^c bcd$ .... Hip.

Hay que demostrar $R^c acd$. Supongamos (cf. Def. 2 (iv)), entonces,

3. $A \to B \in a$ .... Hip.
4. $A \in c$ .... Hip.

hemos de demostrar $B\bar{\in} d$. Pues bien,

5. $A \to B \in b$ .... 1, 3, Lema 2.
6. $B \in d$ .... 4, 5, Def. 2(iv)

P3. $R^c abc \Rightarrow R^c bac$

Prueba. Suponemos [cfr. Def. 2 (iv)]

1. $R^c abc$ .... Hip.
2. $A \to B \in b$ .... Hip.
3. $A \in a$ .... Hip.

4. $\vdash A \rightarrow ((A \rightarrow B) \rightarrow B)$ A2
5. $(A \rightarrow B) \rightarrow B \in a$ . . . . 3, 4 (a es teoría)
6. $B \in c$ . . . . 1, 2, 5, Def. 2(iv)

Pero (6) es lo que teníamos que demostrar a partir de (1), (2) y (3).

P4. $R^2abcd \Rightarrow \exists x(R^c acx$ y $R^c bxd)$

Prueba. Supuesto que hay $a, b, c, d, y \in K^c$ tales que $R^c aby$ $R^c ycd$, hemos de demostrar que hay una teoría $z \in K^c$ tal que $R^c acz$ y $R^c bzd$. Pues bien, dada la hipótesis, definimos un conjunto de fórmulas X como sigue:

$X = \{B | \exists A(A \rightarrow B) \in a$ y $A \in c)\}$

Es decir, X es el conjunto de las fórmulas B tales que $A \rightarrow B$ es fórmula de a y A es fórmula de c. Entonces, se prueba

i) x es teoría.

Prueba. Como en la demostración del Lema 1, el lector puede demostrar con facilidad que x está cerrada por $\mapsto$ y por &.

ii) $R^c acx$.

Prueba. Inmediata a partir de la definición de x y Def. 2 (iv).

iii) $R^c bxd$.

Prueba. (cf. Def. 2(iv)).

1. $A \rightarrow B \in b$ . . . . Hip.
2. $A \in x$ . . . . Hip.
3. $C \rightarrow A \in a (C \in c)$ . . . . 2, Def. de x
4. $\vdash (C \rightarrow A) \rightarrow ((A \rightarrow B) \rightarrow (C \rightarrow B))$ A4
5. $(A \rightarrow B) \rightarrow (C \rightarrow B) \in a$ . . . . 3, 4 (a es teoría)
6. $C \rightarrow B \in y$. . . . . 1, 5 y Def. 2(iv) (pues $R^c aby$ por Hip.)
7. $B \in d$. . . . . 3 ($C \in c$), 6 (pues $R^c ybd$ por Hip.)

Por (i), (ii) y (iii) tenemos una teoría x tal que $R^c acx$ y $R^c bxd$. Indicamos cómo extenderla a una teoría prima z tal que $R^c acz$ y $R^c bzd$. Considérense el conjunto de todas las teorías y tales que $x \subseteq y$ y $R^c byd$. Por el Lema de Zorn hay una teoría máxima z tal que $R^c bzd$. Ahora, $R^c acz$ es inmediato, y que z es prima se demuestra siguiendo una estrategia semejante a la utilizada en la prueba del Lema 2.

P5. $Raaa$.

Prueba. Semejante a la de P2 (Utilícese el teorema de **R** $((A \wedge (A \rightarrow B)) \rightarrow B)$.

P6. $a^{*c*c} = a$.

Prueba. $A \in a^{*c*c}$ syss $\neg A \notin a^{*c}$ syss $\neg \neg A \in a$ syss $A \in a$ (A10, T21; cf. Def. 2 (v), a es teoría).

P7. Si $R^c abc$, entonces $R^c ac^* b^*$.

Prueba. Suponemos (cf. Def. 2(iv))

1. $R^c abc$ . . . . Hip.
2. $A \rightarrow B \in a$ . . . . Hip.
3. $A \in c^*$ . . . . Hip.

4. $B \notin b^*$ .... Hip. (red. ad abs.)
5. $\neg B \in b$ .... 4, Def. 2(v)
6. $\vdash (A \rightarrow B) \rightarrow (\neg B \rightarrow \neg A)$ T23
7. $\neg B \rightarrow \neg A \in a$ .... 2, 6 (a es teoría)
8. $\neg A \in c$ .... 1, 5, 7 (cf. Def. 2(iv))
9. $A \notin c^*$ .... 8, Def. 2(v)

Pero (9) contradice (3). En consecuencia, P7 se cumple y, por tanto, finaliza la prueba del Lema 4: la estructura modelo canónica es, en efecto, una estructura modelo.

Lema 5. Sea $\langle O^c, K^c, R^c, {}^{*c}\rangle$ la estructura modelo canónica. Entonces, hay una función de evaluación canónica v^c tal que para cada fbf A y $a \in K^c$, $v^c(A, a) = 1$ syss $A \in a$.

Prueba. Demostramos a continuación que la función de evaluación canónica es una función de evaluación. Es decir, que v^c cumple con las condiciones (i)-(vi) de Def. 2 (cf. §VI.1). Pues bien, las cláusulas (i)-(iii) son triviales (para (i) utilícese el Lema 2; para (ii) y (iii), los axiomas para $\wedge$ y $\vee$, y las propiedades respecto de estas conectivas de los miembros de K^c). Por tanto, las cláusulas de interés son (iv) y (v).

Cláusula (iv). Subcaso (1.º). Si $v^c(A \rightarrow B, a) = 1$, entonces para cualesquiera $b, c \in K^c$, si $R^c abc$ y $v^c(A, b) = 1$, entonces $v^c(B, c) = 1$.

Prueba. La prueba, muy sencilla, se apoya en la definición v^c y de R^c.

Cláusula (iv). Subcaso (2.º). Si para cualesquiera $b, c \in K^c$, si $R^c abc$ y $v^c(A, b) = 1$, entonces $v^c(B, c) = 1$, entonces $v^c(A \rightarrow B, a) = 1$. Probamos el caso por contraposición. Suponemos, por tanto, $v^c(A \rightarrow B, a) = 0$, y hemos de demostrar que hay $b', c' \in K^c$ tales que $R^c ab'c'$, $v^c(A, b') = 1$ y $v^c(B, c') = 0$.

Pues bien, definimos los conjuntos de fórmulas

$$b = \{C \vdash A \rightarrow C\}$$

y $$c = \{C \mid \exists D(D \in b \text{ y } D \rightarrow C \in a)\}$$

Como en la prueba de los Lemas 1 y 3, el lector puede demostrar fácilmente que b y c son teorías (son conjuntos de fórmulas cerrados por la implicación y la adjunción) y tales que $R^c abc$ (cfr. la definición de c), $A \in b$ ($\vdash A \rightarrow A$) y $B \notin c$ (Si $B \in c$, entonces $v^c(A \rightarrow B, a) = 1$, es decir, $A \rightarrow B \in a$, lo cual es imposible por hipótesis).

Mostramos ahora cómo extender b y c a teorías primas b', c' tales que $R^c ab'c'$, $A \in b'$ y $B \notin c'$. Considérese el conjunto Y de todas las teorías y tales que $c \subseteq y$ y $B \notin y$. Exactamente igual a como se demuestra en el Lema 1, se prueba que hay una teoría prima c' tal que $y \subseteq c'$ y $B \notin c'$. Dado $R^c abc$, es evidente $R^c abc'$. Considérese ahora el conjunto Y de todas las teorías y tales que $b \subseteq y$ y $R^c ayc'$. Exactamente igual a como se demuestra en la prueba del postulado P4, se prueba que hay una teoría prima b' tal que $R^c ab'c'$. Dado $A \in b$, es evidente $A \in b'$. De este modo, tenemos $b', c' \in K^c$ tales que $R^c ab'c'$, $A \in b'$ y $B \notin c'$ como se requería.

Cláusula (v). $v^c(\neg A, a) = 1$ syss $v^c(A, a^*) = 0$

Prueba. Ejercicio para el lector.

Por último, demostramos el

Teorema de Completud: Si $\models A$, entonces $\vdash A$.

Prueba. Demostramos el teorema por contraposición. Supóngase, entonces, $\nvdash A$. Por el Lema 1, hay una teoría T normal y prima tal que $A \notin T$. Por los Lemas 4 y 5, $v^c(A, T) = 0$; es decir, A no es verdadera en el modelo canónico, y, por tanto, $\nvDash A$.

VIII. OTRAS LÓGICAS DE LA RELEVANCIA. SINTAXIS

Disponemos actualmente de innumerables sistemas lógicos de la relevancia. En un trabajo reciente sobre la lógica de la relevancia, L. Peña (1993) ha destacado algunos de ellos clasificándolos en tres apartados: (a) Relevantismo clásico: los sistemas de Anderson y Belnap (cf. *infra*) (b) Relevantismo profundo: los sistemas de la escuela australiana (cf. Routley, Plumwood, Meyer y Brady, 1982) y (c) Relevantismo radical (cf. Avron, 1984; Méndez, 1987a, 1988a, 1988b).

De entre todos estos sistemas destaca, además de **R**, el sistema (clásico) **E** (*entailment*) de la implicación propuesto por Anderson y Belnap como *the true logic*. Los sistemas encuadrables en los otros dos apartados son, en el primer caso (relevantismo profundo) esencialmente restricciones de los sistemas clásicos; en el segundo (relevantismo radical), extensiones del fragmento implicativo de **R** (cf. §IV) complementadas con restricciones del conjunto de axiomas funcionales de verdad de **R.**

Pues bien, los sistemas pertenecientes al apartado (b) son fácilmente accesibles desde el conocimiento de la lógica de la relevancia **R.** Por otro lado, los interesantes sistemas de Avron presentan el inconveniente (desde el punto de vista de la presente exposición) de que su motivación es diferente a la que sustenta a los impulsados por Anderson y Belnap. Hemos optado, pues, por exponer en lo que sigue la fundamentación y estructura de los sistemas **E, RMO, Rm** y **RMOm.**

El sistema **RMO** (cf. Méndez, 1988b) es una alternativa, en la línea de Anderson y Belnap, a **R.** Los sistemas **Rm** y **RMOm** (cf. Méndez, 1987b, 1988b) son, intuitivamente, los fragmentos positivos de **R** y **RMO** con negación mínima al estilo intuicionista, opción ésta descartada habitualmente por los autores relevantistas (pero cf. Tennant, 1987).

La lógica de la implicación ***E***

Anderson y Belnap consideran el sistema **R** como la lógica de la relevancia, la *lógica del condicional relevante.* Ahora bien, ¿es **R** la *lógica de la implicación* en el sentido que Lewis daba al término? Para definir la lógica de la implicación en dicho sentido, Anderson y Belnap, siguiendo precisamente a este autor, exigen que, además de relevancia entre antecedente

y consecuente, se dé entre ellos *relación de necesidad*. Y proponen como traducción formal de este requisito la siguiente:

Propiedad Ackermann. L es un sistema de la *implicación* si $A \rightarrow (B \rightarrow C)$ no es demostrable cuando A no contiene a $\rightarrow$ entre sus símbolos.

Este requisito podría explicarse como sigue. En un sistema de lógica proposicional cuyas únicas conectivas sean $\rightarrow$, $\wedge$, $\vee$ y $\neg$ es la primera de éstas ($\rightarrow$) la única que representa, o podría representar, fórmulas «necesitivas» («necessitive formulae») que, en un sistema estándar de lógica modal, estarían representadas por el operador de necesidad, *L*.

Pues bien, el sistema **E** (descrito en §VIII 1 y §IX 1) es exactamente la restricción de **R** con la propiedad Ackermann.

El sistema ***RMO*** *de la lógica de la relevancia*

Comprobamos en § IV.3.2. que el sistema **R** se construyó definiendo el fragmento implicativo a partir de la siguiente caracterización sintáctica de relevancia:

C4) $A_1 \rightarrow (\ldots \rightarrow (A_n \rightarrow B)\ldots)$ es teorema si *cada* $A_i (1 \leq i \leq n)$ se *usa* en la prueba de B.

Esta caracterización presenta el inconveniente de que en ella no se exige meramente el uso de cada premisa, sino, eventualmente, el *uso de cada aparición de cada premisa*. Por esta razón, algunas tesis como el axioma «mingle»

M. $A \rightarrow (A \rightarrow A)$

no son teoremas de **R**. Este resultado parece, sin duda, contrario a cualquier noción plausible de relevancia entendida sintácticamente como «uso de las premisas en la derivación de la conclusión».

Pues bien, esta dificultad se soluciona modificando C4) como sigue

C4') $A_1 \rightarrow (\ldots \rightarrow (A_n \rightarrow B)\ldots)$ es teorema si *cada* $A_i (1 \leq i \leq n)$ en el conjunto $\{A_1, \ldots, A_n\}$ se *usa* en la prueba de B.

Ahora, **RMO** se construye a partir de C4') como **R** se construyó a partir de C4).

Los sistemas ***Rm*** *y* ***RMOm*** *de la lógica de la relevancia*

Como hemos apuntado más arriba, **Rm** y **RMOm** son el resultado de añadir a **R** y **RMO**, respectivamente, una «negación mínima» al estilo de los cálculos intuicionistas mínimos de Kolmogoroff y Johannsson.

1. *El sistema* ***E*** *de la lógica de la implicación*

i) *Lenguaje formal*. Es el mismo que el del sistema **R**.
ii) *Axiomas*. Son los mismos que los de **R** salvo que la Ley de Aserción

$$A2.\ A \to ((A \to B) \to B)$$

se sustituye por la Ley de Aserción restringida

$$A2'.\ (A \to B) \to (((A \to B) \to C) \to C)$$

iii) *Reglas de derivación.* Además de *Modus ponens* y *Adjunción,* añadimos *Necesitación:* Si $\vdash$ A, entonces $(A \to A) \to A$.

iv) *Teoremas.* El mismo conjunto que el señalado para **R** salvo que la Ley de Permutación de Premisas

$$T3.\ (A \to (B \to C)) \to (B \to (A \to C))$$

es, ahora, en **E** la Ley de Permutación de Premisas restringida

$$T3'.\ (A \to ((B \to C) \to D) \to ((B \to C) \to (A \to D))$$

que se demuestra con A2' y A4 como, en **R,** T3 se demostraba con A2 y A4.

Observación 1. Nótese que *Necesitación* es una consecuencia inmediata de A2 de **R;** como, además, A2' es una restricción de A2, se sigue que **E** es un sistema contenido en **R.** Carece, por tanto, de todos los teoremas clásicos de los que carecía **R,** además de no contar ni con la Ley de Aserción ni con la Ley de Permutación ni con todas las consecuencias de estas dos leyes.

Observación 2. T4 y T5 que en **R** se prueban con T3 pueden probarse ahora en **E** del mismo modo con T3'.

2. *El sistema **RMO** de la lógica de la relevancia*

i) *Lenguaje formal.* El mismo que el de **R.**

ii) *Axiomas.* Todos los de **R** salvo A12. Además añadimos

$$A13.\ A \to (A \to A)$$

iii) *Reglas de derivación.* Las de **R** y, además, la regla *Contraposición* (Si $\vdash A \to B$, entonces $\vdash \neg B \to \neg A$) y *Reductio ad absurdum* (Si $\vdash A \to B$, entonces $\vdash (A \to \neg B) \to \neg A$).

iv) *Teoremas.* T1-T20 de **R** son teoremas de **RMO** y se prueban como en aquel sistema. De los teoremas de **RMO** en los que figura la negación, no son tesis de **RMO,** además de A12, T22, T23, T24 (las leyes de Contraposición) y T26, T27, T28 (las leyes de Reductio). Precisamente, la diferencia entre **R** y **RMO** radicaría, por un lado, en que Contraposición y Reductio sólo valen en **RMO** como reglas de derivación; por otro lado, en la adición del axioma «mingle» A13. El resto de los teoremas de la lista se prueban como en **R** utilizando ahora las reglas Contraposición y Reductio en lugar de los teoremas correspondientes como hacíamos en el caso de *R*.

Observación 1. Tenemos como reglas derivadas todas las variantes de Contraposición y Reductio:

a) Si $\vdash A \to \neg B$, entonces $\vdash B \to \neg A$
b) Si $\vdash \neg A \to B$, entonces $\vdash \neg B \to A$
c) Si $\vdash \neg A \to \neg B$, entonces $\vdash B \to A$

d) Si $\vdash A \rightarrow B$, entonces $\vdash(\neg A \rightarrow B) \rightarrow B$
e) Si $\vdash \neg A \rightarrow \neg B$, entonces $\vdash(\neg A \rightarrow B) \rightarrow A$

Todas ellas se prueban con las reglas primitivas y los teoremas de la doble negación (A10, T21).

Observación 2. T21 se prueba ahora con A1 y Contraposición (b). Por otro lado, A11 no es independiente: se demuestra con A1 y Reductio.

3. *El sistema* **Rm** *La lógica de la relevancia con negación mínima*

i) *Lenguaje formal.* El mismo que **R.**
ii) *Axiomas.* Los mismos que **R** salvo la Ley de Contraposición fuerte

$$\text{A12. } (\neg A \rightarrow B) \rightarrow (\neg B \rightarrow A)$$

no aceptable intuicionistamente, que se sustituye por la Ley de Contraposición débil

$$\text{A12'. } (A \rightarrow B) \rightarrow (\neg B \rightarrow \neg A)$$

que sí es admitida por los intuicionistas. Así se derivan, p. ej.,

$$A \rightarrow \neg\neg A,\ (A \rightarrow \neg A) \rightarrow \neg A,\ (A \rightarrow \neg B)) \rightarrow (B \rightarrow \neg A),$$
$$(A \rightarrow B) \rightarrow (A \rightarrow \neg B) \rightarrow \neg A, \text{ etc.}$$

iii) *Reglas de derivación. Modus ponens* y *Adjunción* como en **R.**
iv) *Teoremas.* T1-T20 (teoremas en los que no interviene la negación) son tesis de **Rm** y se prueban del mismo modo que en **R.** T21-T38 son teoremas en los que interviene la negación; así pues, no son tesis de **Rm** todos los que no son admitidos intuicionistamente. En particular, son teoremas T23, T24, T26, T29, $T\underleftarrow{30}$, $T\underrightarrow{31}$, $T\underrightarrow{32}$, T34, T36 y T38 que se prueban como en **R** (cfr. *Observación 1).* No son, por tanto, teoremas T21, T22, T25, T27, T28, $T\underrightarrow{30}$, $T\underleftarrow{31}$, $T\underleftarrow{32}$, T33, T35 y T37. Por otro lado, tenemos, como en la Lógica intuicionista,

T39. $\neg\neg\neg A \rightarrow A$ A10, A12
T40. $\neg\neg(A \vee \neg A)$ T26, T29

Observación 1. $T\underleftarrow{30}$ designa T30 leído de derecha a izquierda; $T\underrightarrow{31}$ designa T31 leído de izquierda a derecha.

Observación 2. T23 es A12. $T\underrightarrow{31}$ se prueba en **Rm** con A5, A8 y T24; $T\underrightarrow{32}$ con A5, A8 y A12'. En la prueba de T40, derívese previamente $(A \rightarrow (B \wedge \neg B)) \rightarrow \neg A$ con T16, T17 y T26. El resto de los teoremas se prueba como en **R.**

4. *El sistema* ***RMOm:*** *la lógica de la relevancia* ***RMO*** *con negación mínima*

i) *Lenguaje formal.* El mismo que **R.**
ii) *Axiomas.* Todos los de **R** salvo A12 que se sustituye por

$$\text{A12'. } (A \rightarrow B) \rightarrow (\neg B \rightarrow \neg A)$$

Además, añadimos

$$\text{A13. } A \rightarrow (A \rightarrow A)$$

iii) *Reglas de derivación. Modus ponens* y *Adjunción* como en **R.**
iv) *Teoremas.* Como se deduce por la descripción anterior, **RMOm** es el resultado de añadir A13 a **Rm.** Así pues, el conjunto de teoremas de **RMOm** es el mismo que el de **Rm** más todas las consecuencias de A13 entre las que se encuentra, p. ej., *la Conversa de la Ley de Contracción* $(A \rightarrow B) \rightarrow (A \rightarrow (A \rightarrow B))$ que, por cierto, es intercambiable con A13 como axioma en este sistema y en **RMO.**

IX OTRAS LÓGICAS DE LA RELEVANCIA. SEMÁNTICA

1. *Semántica para **E***

i) *Estructura modelo.* Una Estructura modelo para **E** es una estructura del tipo $\langle O, K, R, {}^* \rangle$ donde $O, K, R, {}^*$ y d1, d2, P1, P2, P4, P5, P6 y P7 son como en las estructuras modelo **R.** La única diferencia, por tanto, estriba en que P3 se sustituye por

P3'. $\exists x(x \in O$ y $Raxa)$

ii) *Modelos. Validez.* Un modelo para **E** es una estructura del tipo $\langle O, K, R, {}^*, \text{v} \rangle$ donde $\langle O, K, R, \text{v} \rangle$ es una estructura modelo y v cumple con las cláusulas (i)-(vi) como en los modelos para **R.** La definición de validez es la misma.
iii) *Consistencia. Completud.* Para probar los Teoremas de Consistencia y Completud seguimos la misma estrategia que en el caso de **R.** Dadas las diferencias entre ambos sistemas, para probar la consistencia de **E** hay que demostrar que A2' y Necesitación son válidos: utilícese P3' en ambos casos. Para probar la completud de **E** sólo hay que demostrar que P3' se cumple cuando se define canónicamente (cfr. Lema 4). Pues bien, esto puede hacerse como sigue. Defínase $X = \{A \mid (A \rightarrow A) \rightarrow A\}$. Es, entonces, fácil demostrar que x es una teoría normal tal que $R^c axa$. Es decir, $\exists x(x \in O^c$ y $R^c axa)$. Después, x se extiende a una teoría prima z tal que $z \in O^c$ y $R^c aza$ siguiendo el procedimiento habitual.

2. *Semántica para **RMO***

i) *Estructuras modelo.* Una estructura modelo para **RMO** es una estructura del tipo $\langle O, K, R, {}^* \rangle$ donde $O, K, R, {}^*$ y d1, d2, P1-P6 son como en las estructuras modelo para **R.** Se diferencian de ellas en que P7 se sustituye por

P7'. Si $a \leq b$, entonces $b^* \leq a^*$

y en que se añaden los postulados

P8. Si $Rabc$, entonces $a \leq c$ ó $b \leq c$

P9. Ra^*aa^* (o Raa^*a)

ii) *Modelos. Validez.* Un modelo para **RMO** es una estructura del tipo $\langle O, K, R, {}^*, \text{v} \rangle$ donde $\langle O, K, R, {}^* \rangle$ es una estructura modelo

y v cumple con las cláusulas (i)-(vi) como en el caso de los modelos para **R.** La definición de validez es la misma.

iii) *Consistencia. Completud.* Para demostrar el Teorema de Consistencia sólo hace falta demostrar que A13, Contraposición y Reductio son válidas: utilícense P8', P7' y P9, respectivamente. Para probar que P7', P8 y P9 se cumplen cuando se interpretan canónicamente (cfr. Lema 4), utilícense, con la ayuda del Lema 2, Contraposición, A13 y Reductio, respectivamente. Hecho esto, es evidente que entonces se sigue el Teorema de Completud.

3. *Semántica para* ***Rm***

i) *Estructuras modelo.* Una estructura modelo para **Rm** es una estructura del tipo ⟨*O, K, R*⟩ donde *O, K, R,* d1, d2, y P1-P5 son como en las estructuras modelo para **R.**

ii) *Modelos. Validez.* Un modelo para **Rm** es una estructura del tipo ⟨*O, K, R,* v⟩ donde ⟨*O, K, R*⟩ es una estructura modelo, v cumple con las condiciones (i)-(v) de los modelos para **R,** y, finalmente, la cláusula (vi) se sustituye por
(vi') $v(\neg A, a) = 1$ syss para cualquier $b \in K$ y $c \in K - S$, $v(A, b) = 0$ ó *no-Rabc.*
(*S* es un subconjunto cualquiera de *K*).
La definición de validez es la misma.

iii) *Consistencia. Completud.* Dado que la cláusula de interpretación de la negación es nueva, para demostrar el Teorema de Consistencia hay que probar que A10, A11 y A12' son válidos: utilizar P4, P5(b) (cfr. §VI.2) y P5, respectivamente. Para demostrar el Teorema de Completud hay que probar que la cláusula (vi') se cumple canónicamente: sígase una estrategia semejante a la empleada en la prueba de la cláusula (v).

4. *Semántica para* ***RMOm***

i) *Estructura modelo.* En todo iguales a las definidas por Rm salvo por el hecho de que se añade el postulado (cfr. §IX.2)
P8. Si *Rabc,* entonces $a \leq c$ ó $b \leq c$

ii) *Modelos. Validez.* Se definen, con la salvedad anotada en el parágrafo anterior, exactamente igual que los modelos para **Rm.**

iii) *Consistencia. Completud.* Es evidente que para demostrar el Teorema de Consistencia sólo hay que probar que A13 es válido. Como en el caso de **RMO,** utilícese P8. Para probar el Teorema de Completud sólo hay que demostrar que P8 se cumple cuando se interpreta canónicamente (el resto de la prueba es como la correspondiente para **Rm**). Pues bien, al igual que hicimos en el caso de **RMO,** utilícense el Lema 2 y A13.

X. OTROS RESULTADOS. CONCLUSIONES

Mencionamos a continuación algunos resultados capitales sobre el tema que, debido a su complejidad, no hemos podido exponer en este artículo. Finalizamos con algunas conclusiones sobre lo expuesto en este trabajo.

Otros resultados

1. Indecidibilidad de toda lógica comprendida entre $\mathbf{T}_+$**-W** y **K.R.** A. Urquhart (1984) (cf. también Anderson, Belnap y Dunn, 1992) ha demostrado que cualquier sistema de lógica comprendido entre $\mathbf{T}_+$**-W** y **K.R.** es indecidible. El sistema $\mathbf{T}_+$**-W** es el fragmento positivo de **R** sin A3; **K.R.** es el sistema **R** más el axioma $(A\wedge\neg A)\rightarrow B$. Entre los sistemas encuadrables en el espectro definido por $\mathbf{T}_+$**-W** y **K.R.** están **R** y **E.** No está, sin embargo, **RMO.**
2. Incompletud de la semántica cuantificacional de dominio único o constante:
   K. Fine (1989) (cf. también Anderson, Belnap y Dunn, 1992) ha probado que los sistemas estándar de lógica de la relevancia cuantificacionales son incompletos respecto de la semántica habitual cuando el dominio en todos los mundos posibles es único o constante.
3. Completud de la lógica de la relevancia respecto de modelos con dominio variable:
   K. Fine (cf. Fine, 1988; también Anderson, Belnap y Dunn, 1992) ha demostrado la completud de los sistemas estándar de lógica de la relevancia (entre ellos, **R** y **E**) respecto de la semántica operacional (cf. Fine, 1974) cuantificacional con un dominio de individuos arbitrarios o genéricos (cf. también Fine, 1985).
4. Aplicaciones de la lógica de la relevancia a la informática:
   Cf. Anderson, Belnap y Dunn, 1992, §VIII.3.
5. Resultados sobre lógicas de la relevancia de orden superior y sobre Aritmética relevante:
   Cf. Anderson, Belnap y Dunn, 1992, cap. XI.

Conclusiones

1. En contra de lo aventurado por Lewis, la definición de una lógica no paradójica no es tarea imposible; más aún, las lógicas no paradójicas son lo suficientemente potentes como para que sean dignas de consideración.
2. A pesar de que el sistema de la lógica de la relevancia **R** (y los que han tratado en §VIII y §IX) están suficientemente definidos desde el punto de vista formal, es preciso reconocer que su interpretación intuitiva es menos clara como así hemos comprobado al mostrar la insuficiencia tanto de la caracterización sintáctica como de la semántica.

3. Precisamente este último hecho explica que la lógica de la relevancia sea *esencialmente relativa:* hay muchas posibilidades de seleccionar un *subconjunto* máximo posible de reglas que no den lugar a paradojas (cf. §IV.3). Algunas de estas posibilidades han sido comentadas en §VIII y §IX.
4. La relatividad de la lógica de la relevancia debe causar tan poco (o tanto) escándalo como la relatividad de la lógica modal.
5. La lógica de la relevancia, cuyo origen está en el artículo de W. Ackermann (1956) o quizá, como aquí hemos defendido, en los trabajos de Lewis (Lewis y Langford, 1932) es, por todo lo expuesto, un campo abierto. Falta, sobre todo, creemos, una explicación convincente de la relación entre los sistemas formales (y su semántica) y las nociones intuitivas que los avalan; falta la definición de un marco general en el que todos los sistemas actuales sean encuadrables; falta desarrollar en profundidad las aplicaciones informáticas, los resultados sobre las correspondientes lógicas de orden superior, y, así, un largo etcétera.

BIBLIOGRAFÍA

En Anderson, Belnap y Dunn (1992) se incluye una bibliografía exhaustiva sobre Lógica de la relevancia hasta aproximadamente 1989. Por esta razón, en lo que sigue sólo incluimos:

a) Trabajos históricos fundamentales en el desarrolo de la Lógica de la relevancia.
b) Trabajos que inciden directamente sobre alguno de los apartados en los que hemos subdividido este artículo.
c) Trabajos cuya perspectiva es esencialmente filosófica, y
d) Trabajos de autores iberoamericanos sobre el tema.

Ackermann, W. (1956), «Begründung einer strengen Implikation»: *The Journal of Symbolic Logic,*
ci
México, 1980).

Anderson, A. R. y Belnap, N. D., Jr. (1962), «The pure calculus of entailment»: *The Journal of Symbolic*

Anderson, A. R y Belnap, N. D., Jr. (1975), *Entail*
sitr, I, Princeton University Press, Princeton.

Anderson, A. R., Belnap, N. D., Jr. y Dunn, J. M. (1992), *Entail*
vance and necessity,

Avron, A. (1984), «Relevant entailment-semantics and formal systems»: *The Journal of Symbolic*

Belnap, N. D., Jr. (1981), «Modal and relevance logics: 1977», en E. Agazzi (ed.) *Modern Logic.*

Belnap, N. D., Jr. y Dunn, J. M. (1981), «Entailment and the disjunctive syllogism», en G. Flöistad y G. H. von Wright (eds.), *Philosophy*
Martinus Nijhoff, Den Kogueris, La Haya 337-366.

Burgess, J. P. (1981), «Relevance: a fallacy?»: *Notre Dame Journal of formal logic,* 97-104.

Burgess, J. P. (1983), «Common sense and relevance»: *Notre Dame Journal of formal logic, 24,* 41-53.

Burgess, J. P. (1984), «Read on relevance: a rejoinder»: *Notre Dame Journal of formal logic, 25,* 217-223.
Church, A. (1951), «The weak theory of implication», en A. Menne, A. Wilhelmy y H. Angsil (eds.), *Kontrolliertes Denken, Untersuchungen zum Logikkalkül und zur Logik der Einzelwissenschaften,* Kommissions-Verlag Karl Alber, München, 22-37.
Copeland, B. J. (1979), «On when a semantics is not a semantics: some reasons for disliking the Routley-Meyer semantics for relevance logic»: *Journal of philosophical logic, 8,* 399-413.
Copeland, B. J. (1982), «The trouble Anderson and Belnap have with relevance»: *Philosophical studies,* vol. 37, pp. 325-333.
Copeland, B. J. (1983a), «Pure semantics and applied semantics. A response to Routley, Routley, Meyer and Martin "On the philosophical bases of relevant logics semantics"»: *Topoi, 22,* 192-204.
Copeland, B. J. (1983b), «A rejoinder to Routley, Meyer and Martin»: *The Journal of non-classical logic, 22,* 61-66.
Curley, E. M. (1975), «The development of Lewis' theory of strict implication»: *Notre Dame Journal of formal logic, 16,* 517-527.
Da Costa, N. C. A. (1974), «On the theory of inconsistent formal systems»: *Notre Dame Journal of formal logic, 15,* 497-510.
Dunn, J. M. (1979), «R-mingle and beneath: extensions of the Routley-Meyer semantics for R»: *Notre Dame Journal of formal logic, 20,* 369-376.
Dunn, J. M. (1986), «Relevance logic and entailment», en Gabbay y Guenthner, 1986, 117-129.
Fine, K. (1974), «Models for entailment»: *Journal of philosophical logic, 3,* 347-372.
Fine, K. (1985), *Reasoning with arbitrary objects*
Fine, K. (1988), «Semantics for quantified relevance logic»: *Journal of philosophical logic, 17,* 27-59.
Fine, K. (1989), «Incompleteness for quantified relevance logics», en Norman y Sylvan, 1989, 205-225.
Font, J. M. y Rodríguez, G. (1990), «Note on algebraic models for relevance logic»: *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik, 146,* 535-540.
Gabbay, D. V. y Guenthner, F. (1986), *Handbook of Philosophical Logic,* Reidel, Dordrecht.
Hacking, I. (1963), «What is strict implication?»: *The Journal of Symbolic Logic, 28,* 51-71.
Leblanc, H. (1973), *Truth, Syntax and modality,* North Holland, Amsterdam.
Lewis, C. I. y Langford, C. H. (1932), *Symbolic Logic,* Dover Publications, New York, [2]1959.
Méndez, J. M. (1986), «Una crítica inmanente de la lógica de la relevancia»: *Crítica, 18,* 61-94.
Méndez, J. M. (1987a), «A Routley-Meyer semantics for Converse Ackermann Property»: *Journal of philosophical logic, 16,* 65-76.
Méndez, J. M. (1987b), «Constructive R»: *Bulletin of the Section of Logic, vol.* 167-175.
Méndez, J. M. (1988a), «Converse Ackermann Property and semiclassical negation»: *Studia Logica, 47,* 159-163.
Méndez, J. M. (1988b), «The compatibility of relevance and mingle»: *Journal of philosophical logic, 17,* 279-297.
Méndez, J. M. (1989), «E-Mingle has the variable sharing property» (resumen): *The Journal of Symbolic Logic, 51,* 663-664.
Méndez, J. M. (1991), «Introducción a los conceptos fundamentales de la lógica de la relevancia»: *Arbor, 520,* 75-93.
Meyer, R. K. (1973), «Intuitionism, entailment, negation», en Leblanc, 1973, 168-198.

Meyer, R. K. (1976), «Metacompleteness»: *Notre Dame Journal of formal logic, 17,* 501-516.
Meyer, R. K. y McRobbie, M. A. (1982), «Multisets and relevant implication»: *Australasian Journal of Philosophy, 60,* 265-281.
Meyer, R. K. y Mortensen, C. (1984), «Inconsistent models for relevant arithmetic»: *The Journal of Symbolic Logic, 49,* 917-929.
Morado, J. R. (1984), *¿Hay rivales para la lógica clásica? El caso de las lógicas relevantes y las lógicas libres,* Universidad de México.
Mortensen, C. (1983), «The validity of the disjunctive syllogism is not so easily proved»: *Notre Dame Journal of formal logic, 24,* 35-40.
Mortensen, C. (1086), «Reply to Burgess and to Read»: *Notre Dame Journal of formal logic, 27,* 195-200.
Norman, J. y Sylvan, R. (1989), *Directions in relevant logic,* Kluwer, Dordrecht.
Orayen, R. (1983), «Deducibility implies relevance? A negative answer»: *Crítica, vol. 15, 43,* 3-29; *44,* 3-25.
Orayen, R. (1986), «Una evaluación de las críticas relevantistas a la lógica clásica», en S. Álvarez, F. Broncano y M. A. Quintanilla (eds.), *Actas del I Simposio Hispano-Mexicano de Filosofía,* Edic. Univ. de Salamanca, Salamanca.
Peña, L. (1984), «Tres enfoques en lógica paraconsistente»: *Contextos, 3,* 81-130.
Peña, L. (1993), «Is reasoning the same as relevant inference?», manuscrito.
Read, S. (1983), «Burgess on relevance: a fallacy indeed»: *Notre Dame Journal of formal logic, 24,* 437-481.
Routley, R. y Meyer, R. K. (1972a), «The semantics of entailment. II»: *Journal of philosophical logic.*
Routley, R. y Meyer, R. K. (1972b), «The semantics of entailment. III»: *Journal of philosophical logic, 1,* 192-208.
Routley, R. y Meyer, R. K. (1973), «The semantics of entailment. I», en Leblanc, 1973, 199-243.
Routley, R. y Meyer, R. K. (1982), «The semantics of entailment. IV», en Routley, Plumwood, Meyer y Brady, 1982, 407-424.
Routley, R., Plumwood, V., Meyer, R. K. y Brady, R. T. (1982), *Relevant logics and their rivals,* Ridgeview Publishing Co., Atascadero, California.
Routley, R., Routley, V., Meyer, R. K. y Martin, E. P. (1982), «On the philosophical bases of relevant logic semantics»: *The Journal of non-classical logic, 1,* 71-105.
Scott, D. (1971), «On engendering an illusion of understanding»: *The Journal of philosophy, 68,* 787-807.
Sylvan, R. (antes R. Routley) (1988), «A preliminary western history of sociative logics», manuscrito.
Tennant, N. (1987), *Antirealism and logic,* Clarendon Press, Oxford.
Urquhart, A. (1972), «Semantics for relevant logics»: *The Journal of Symbolic Logic, 37,* 159-169.
Urquhart, A. (1984), «The undecidability of entailment and relevant implication»: *The Journal of Symbolic Logic, 49,* 1.059-1.073.
Urquhart, A. (1989), «What is relevant implication?», en Norman y Routley, 1989.

COMPUTABILIDAD

Jesús Mosterín

I. PROBLEMAS Y ALGORITMOS

La aparición y desarrollo de los métodos recursivos (o algoritmos) es quizás una manifestación de la llamada ley del mínimo esfuerzo. En efecto, tras realizar un cierto esfuerzo intelectual para resolver un problema de un determinado tipo, y volver a realizar una y otra vez un nuevo esfuerzo intelectual para resolver otros problemas del mismo tipo, se nos puede ocurrir la idea de ahorrarnos en el futuro ese tipo de esfuerzos mediante la invención de un procedimiento de resolución automática de todos los problemas del tipo dado. Es cierto que para inventar tal procedimiento tendríamos que espabilarnos considerablemente y hacer un esfuerzo intelectual notable, pero a partir de ese momento no volveríamos a preocuparnos por los problemas de esa clase: cada vez que se nos plantease uno, podríamos resolverlo automáticamente, sin pensar ni realizar esfuerzo intelectual alguno, simplemente siguiendo las instrucciones del procedimiento al pie de la letra. O, alternativamente, podríamos programar un computador con las instrucciones de nuestro método y dejar que fuera el computador el que fuese resolviendo los problemas. Claro que para ello sería necesario que la aplicación del método no requiriese iniciativa, creatividad, imaginación alguna, sino que bastase con la ciega aplicación de unas instrucciones unívocas y «mecánicas». Precisamente son ese tipo de métodos los que reciben el nombre de métodos recursivos o algoritmos.

Cuando hay un algoritmo para solucionar mecánicamente todos los problemas de una determinada clase, decimos que esa clase de problemas es algorítmicamente dominable.

La tarea de la teoría de la recursión consiste en: 1) precisar el concepto de algoritmo o método recursivo; 2) determinar qué clases de problemas son algorítmicamente dominables y cuáles no lo son; y 3) ofrecer

algoritmos (o métodos recursivos) para la resolución automática de los problemas de cada clase algorítmicamente dominable de problemas.

Desde un punto de vista lógico nos interesan especialmente los problemas que se plantean a nivel lingüístico (en el sentido más amplio de esta expresión), es decir, problemas relacionados con filas de signos o expresiones sobre un alfabeto determinado. Estos problemas se dividen en tres grupos principales: problemas de computación, problemas de decisión y problemas de generación.

Problemas de computación son aquéllos en que se nos pide hallar el valor de una determinada función para un (o para varios, si la función es de varios argumentos) determinado argumento. Así, por ejemplo, los problemas de «cálculo» que se plantean en las escuelas —«calcúlese la raíz cuadrada de 820, el triplo de 17, el máximo común divisor de 10, 12 y 14, el mínimo común múltiplo de 4, 7 y 132, el producto de 8 por 375», etc.— son todos problemas de computación. También los problemas de traducción son problemas de computación, suponiendo una función que a cada oración de una lengua aplica unívocamente otra oración de una segunda lengua, aquélla cuyo significado más corresponde al de la primera.

Problemas de decisión son aquéllos en que se nos pide que averigüemos si una determinada expresión tiene una cierta propiedad (o pertenece a un cierto conjunto) o si varias expresiones están en una determinada relación. Así, por ejemplo, el problema de averiguar si un determinado número natural es primo o no, es un problema de decisión. También son problemas de decisión el problema de averiguar (o decicir) si una determinada fórmula de la lógica sentencial es una tautología o no, o si una determinada fórmula de la lógica de primer orden es (lógicamente) válida o no, o si una determinada fila de signos del alfabeto de nuestra lengua constituye una oración castellana gramaticalmente correcta o no, etc.

Problemas de generación son aquéllos en los que se nos pide que escribamos o generemos sucesivamente todas las expresiones de un determinado conjunto. Así, por ejemplo, el problema de generar sucesivamente todas las fórmulas válidas de la lógica de primer orden (al que responde la creación de los cálculos deductivos) es un problema de generación. La gramática generativa de una lengua natural trata —como su nombre indica— de resolver un problema de generación: el de generar sucesivamente todas y solas las filas de signos que constituyen oraciones gramaticalmente correctas de esa lengua.

II. CONCEPTOS RECURSIVOS

Supongamos en lo sucesivo que hablamos siempre de las expresiones (o filas de signos) formables con los signos de un determinado alfabeto finito. En especial, cuando hablemos de números naturales nos referiremos a

las expresiones sobre el alfabeto {I}: I, II, III, IIII, IIIII, etc. Cuando digamos «conjunto» queremos decir «conjunto de expresiones», etc.

Una función es recursivamente computable si y sólo si hay un algoritmo que para cada uno de sus argumentos x (o cada n de sus argumentos $x_1,\ldots,x_n$, si es n-ádica) nos permite obtener en un número finito de pasos su correspondiente valor f(x) (o, si es n-ádica, $f(x_1,\ldots,x_n)$).

Un conjunto A es recursivamente decidible si y sólo si hay un algoritmo que para cada expresión x nos permite averiguar en un número finito de pasos si $x \in A$ o $x \notin A$. (Una relación n-ádica R es recursivamente decidible si y sólo si hay un algoritmo que para cada n expresiones $x_1,\ldots,x_n$ nos permite decidir en un número finito de pasos si $\langle x_1 \ldots x_n \rangle \in R$ o no).

Dado un conjunto cualquiera A, llamamos función característica de A al función J_A tal que:

$J_A(x) = 0$ si $x \in A$
$J_A(x) = 1$ si $x \notin A$.

Igualmente llamamos función característica de una relación R a la función J_R tal que:

$J_R(x_1,\ldots,x_n) = 0$ si $\langle x_1,\ldots,x_n \rangle \in R$
$J_R(x_1,\ldots,x_n) = 1$ si $\langle x_1,\ldots,x_n \rangle \notin R$.

Ahora bien, está claro que el conjunto A (o la relación R) es decidible si y sólo si su correspondiente función característica J_A (o J_R) es computable. Supongamos que A es decidible. Entonces J_A será computable. En efecto, dado un argumento cualquiera x, el valor de J_A para x será 0 ó 1 según que $x \in A$ o no, lo que hemos supuesto que es decidible. Supongamos a la inversa que J_A es computable. Entonces A será decidible. En efecto, dado un x cualquiera bastará computar ($J_A(x)$ y ver si el resultado es 0 (en cuyo caso sabremos que $x \in A$) o 1 (y entonces $x \notin A$). Así, pues, vemos que la decibilidad de conjuntos (o relaciones) es reducible a la computabilidad de funciones.

Un conjunto A es (recursivamente) generable si y sólo si hay un algoritmo para producir sucesivamente todas y solas las expresiones de A. Este concepto de generabilidad recursiva de un conjunto es también reducible al de computabilidad de una función. Como fácilmente se ve, un conjunto A es (recursivamente) generable si y sólo si hay una función computable cuyo dominio es el conjunto ω de los números naturales y cuyo contradominio es A. Por eso en vez de hablar de conjuntos generables se habla a veces de conjuntos recursivamente numerables.

Vemos, pues, que los conceptos de conjunto o relación decidible y de conjunto generable son reducibles al de función computable. Y ya hemos dicho que una función es computable si hay un algoritmo para computar el valor de cada uno de sus argumentos. Pero ¿qué es un algoritmo? Un conjunto de instrucciones completamente unívocas cuyo seguimiento y aplicación no requiere iniciativa alguna. A primera vista, podría

pensarse que el concepto de algoritmo está claro y no requiere mayores precisiones. Podría pensarse que es fácil decidir en cada caso si un conjunto dado de instrucciones, si un texto determinado que se nos presenta como algoritmo, es efectivamente un algoritmo o no. Pues bien, no es ése el caso. La propiedad de ser un algoritmo (o, si se prefiere, el conjunto de los algoritmos respecto al de los textos) es indecible. Incluso el concepto —más simple— de algoritmo de computación es indecidible, como a continuación mostramos.

III. INDECIDIBILIDAD DEL CONCEPTO DE ALGORITMO

Supongamos que los algoritmos —series finitas de instrucciones— estén formulados en castellano. Un algoritmo tendrá siempre una longitud determinada (y finita). Supongamos que hemos ordenado los algoritmos de computación por su longitud, es decir, por el número de letras (incluyendo entre éstas los espacios en blanco, las cifras y los signos de puntuación) que contengan (primero los más cortos, luego, los que tienen un signo más, etc.) y, dentro de la misma longitud, lexicográficamente (es decir, primero los que empiezan por «a», luego los que empiezan por «b», etc., como en un diccionario en el que un algoritmo entero se considerase como una sola palabra). Sea A_1, A_2, A_3, A_4... la sucesión así ordenada de todos los algoritmos de computación.

Una función es computable si y sólo si hay un algoritmo para computarla (es decir, para computar su valor para cada uno de sus argumentos). A la función computable mediante el algoritmo A_n la llamaremos f_n.

Está claro que una función cualquiera f es computable si y sólo si hay algún número natural n tal que A_n computa f, es decir, tal que $f = f_n$. Toda función computable es una f_n.

Ahora definimos la función h para todos los números naturales n de la siguiente manera:

$$h(n) = f_n(n) + 1.$$

Si h fuera una función computable, coincidiría con alguna de las f_i para algún número natural i. Pero, para cualquier i, h discrepa de f_i en el valor que aplica a i. En efecto,

$h \neq f_1$, pues $h(1) = f_1(1) + 1 \neq f_1(1)$
$h \neq f_2$, pues $h(2) = f_2(2) + 1 \neq f_2(2)$
$h \neq f_3$, pues $h(3) = f_3(3) + 1 \neq f_3(3)$
....
....
....
$h \neq f_i$, pues $h(i) = f_i(i) + 1 \neq f_i(i)$.

Por tanto, h no es una función computable.

Ahora bien, si el conjunto de los algoritmos de computación fuese efectivamente generable, h sería computable. En efecto, para cada número natural n, para computar h(n) bastaría con generar A_1, A_2, A_3,... hasta A_n, computar $f_n(n)$ con ayuda de A_n y añadir 1 al resultado. Así obtendríamos h(n). Pero ya hemos visto que h no es computable. Por tanto, el conjunto de los algoritmos de computación no es generable.

Si la propiedad de ser un algoritmo (o, si se prefiere, si el conjunto de los algoritmos) fuese decidible respecto al conjunto de los textos, podríamos generar el conjunto de los algoritmos del siguiente modo: Empezaríamos por generar todos los posibles textos castellanos (con sentido o sin él). Esto lo conseguiríamos generando sucesivamente todas las combinaciones con repetición posibles de n signos del alfabeto castellano extendido (que abarca letras, cifras, espacios de separación y signos de puntuación); primero, las de un signo, luego, las de dos, las de tres, etc. Según fuéramos generando los textos, iríamos decidiendo para cada uno de ellos si se trata de un algoritmo de computación o no, e iríamos apuntando en lista aparte todos los textos que efectivamente fuesen algoritmos de computación. Así habríamos obtenido un procedimiento para generar todos los algoritmos de computación. Pero esto es imposible, pues ya hemos visto que el conjunto de los algoritmos de computación no es generable. Por tanto, la propiedad de ser un algoritmo (de computación) no es decidible respecto al conjunto de los textos. Es decir, no hay (no puede haber) un algoritmo con cuya ayuda podamos decidir de cada presunto algoritmo si es realmente un algoritmo o no. Con lo que queda suficientemente probada la no-trivialidad y la necesidad de dilucidación del concepto de algoritmo (de computación).

IV. LA PRECISIÓN DE TURING

Se llaman funciones recursivas (o recursivamente computables) aquellas funciones para cuya computación hay un algoritmo. Se han propuesto diversas dilucidaciones o precisiones de este concepto, pero todas han resultado equivalentes. El primero que propuso identificar el concepto intuitivo de algoritmo con un concepto precisado fue A. Church, en 1936 («An unsolvable problem of elementary number theory»). De entre las diversas precisiones equivalentes del concepto de algoritmo elegimos aquí la de A. M. Turing, también de 1936 («On computable numbers with an application to the Entscheidungsproblem»).

Turing introdujo el concepto de máquina de Turing. No se trata, claro está, de ninguna máquina física, real, sino de un esquema abstracto para representar conjuntos de instrucciones unívocas; algo así como el programa de un computador. La precisión de Turing consiste en proponer que consideremos que una función es computable si y sólo si hay una máquina de Turing para computarla, que un conjunto o relación es decidible si y sólo si hay una máquina de Turing para decidirla, y que un

conjunto es generable si y sólo si hay una máquina de Turing para generarlo. Puesto que ya vimos que los conceptos de decidibilidad y generabilidad son reducibles al de computabilidad, nos limitaremos aquí a este último. Y puesto que Gödel mostró cómo los problemas de computación de funciones sobre expresiones de alfabetos finitos pueden traducirse a problemas de computación de funciones numéricas, nos limitaremos también a estas últimas.

Los números naturales son aquí identificados con las filas de palotes: I, II, III, etc., es decir, con las expresiones sobre el alfabeto {I}. Una función numérica f aplica a cada número natural n (que aquí representaremos mediante n + 1 palotes, a fin de que el palote sólo represente al 0) otro número natural f(n) (que aquí representaremos mediante f(n) + 1 palotes). Un algoritmo de computación para esa función nos indicará cómo hemos de manipular la expresión de partida (el argumento) para llegar, en un número finito de pasos y siguiendo al pie de la letra las instrucciones, hasta la expresión de llegada (el valor). Las instrucciones son tan precisas que hasta una máquina podría seguirlas. Pues bien, imaginémonos una máquina que trabaja una cinta compuesta de cuadros. En cada cuadro sólo puede haber o un palote, I, o nada, *. En un momento dado la máquina ve un solo cuadro de la cinta (cuadro de trabajo) y se encuentra en un estado determinado. El estado de la máquina y la inscripción del cuadro de trabajo determinan unívocamente el siguiente paso de la máquina, que consiste necesariamente en una de estas cinco cosas: marcar un palote en el cuadro de trabajo (I), marcar el signo vacío (es decir, borrar) en el cuadro de trabajo (*), pasar al cuadro de la derecha (r), pasar al cuadro de la izquierda (l) o pararse (S). El programa de esa máquina consistirá en la indicación de qué es lo que hará la máquina en cada uno de sus estados, tanto si ve I como si ve * en el cuadro de trabajo, y a qué estado pasará a continuación. Y, como hemos dicho antes, la máquina de Turing no es nada físico, sino que se identifica con su propio programa.

He aquí la definición precisa:

Una máquina de Turing M (sobre el alfabeto {I}) es una tabla o matriz de 4 columnas y 2m filas de la siguiente forma:

1	*	P_{11}	C_{11}
1	I	P_{12}	C_{12}
2	*	P_{21}	C_{21}
2	I	P_{22}	C_{22}
.	.	.	.
.	.	.	.
.	.	.	.
m	*	P_{m1}	C_{m1}
m	I	P_{m2}	C_{m2}

donde para cada i, j($1 \leq i \leq m$, $j = 1$ o $j = 2$):$P_{ij} \in \{*, I, r, l, S\}$
donde para cada i, j($1 \leq i \leq m$, $j = 1$ o $j = 2$):$C_{ij} \in \{1, 2, \ldots, m\}$

Fijémonos en una fila cualquiera de la tabla. El primer signo indica un estado en que se puede encontrar la máquina. El segundo, una inscripción posible del cuadro de trabajo. El tercero, el paso que deberá dar la máquina cuando, encontrándose en dicho estado, vea tal inscripción en el cuadro de trabajo. El cuarto, el estado en el que se encontrará la máquina después de haber dado ese paso. Los números, 1, 2,... hasta m corresponden a los distintos estados «internos» de la máquina (recuérdese que todo esto es una metáfora). El 1, al estado inicial.

A continuación presentamos una serie de máquinas de Turing que realizan determinadas tareas. En primer lugar, las máquinas elementales: I (que escribe un palote y se para), * (que borra y se para), r (que da un paso a la derecha —es decir, el cuadro inmediato a la derecha se convierte en nuevo cuadro de trabajo— y se para) y l (que da un paso a la izquierda y se para). He aquí sus tablas:

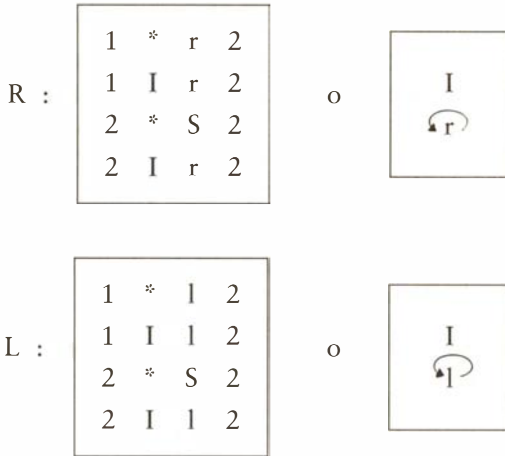
I				*				r				1			
1	*	I	2	1	*	*	2	1	*	r	2	1	*	l	2
1	I	I	2	1	I	*	2	1	I	r	2	1	I	l	2
2	*	S	2	2	*	S	2	2	*	S	2	2	*	S	2
2	I	S	2	2	I	S	2	2	I	S	2	2	I	S	2

Veamos ahora la máquina R (que va hasta el primer cuadro vacío a la derecha de una expresión o fila de palotes dada, y allí se para) y la máquina L (que hace lo propio, pero hacia la izquierda). Representando por W una expresión (fila de palotes) cualquiera, por * un cuadro vacío y por ~ un cuadro cualquiera, poniendo encima de la raya la situación de partida de la cinta y debajo, la de llegada, y señalando mediante una flechita inferior el cuadro de trabajo, podemos describir la acción de R y L así:

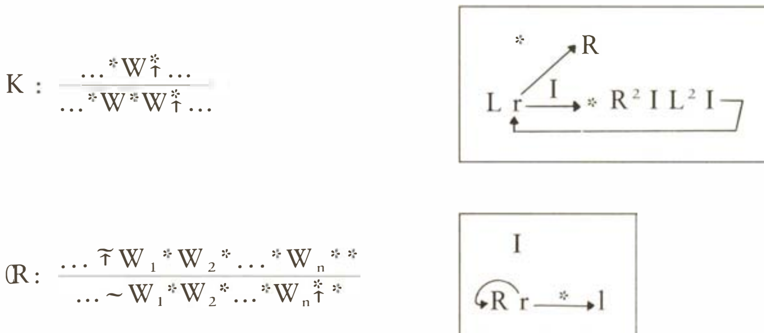
$$R : \frac{\ldots \underset{\uparrow}{\sim} W * \ldots}{\ldots \sim W \underset{\uparrow}{*} \ldots}$$

$$L : \frac{\ldots * W \underset{\uparrow}{\sim} \ldots}{\ldots \underset{\uparrow}{*} W \sim \ldots}$$

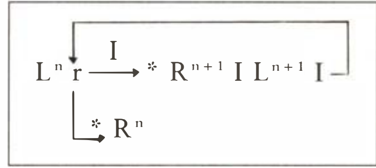
Una máquina de Turing puede representarse mediante su tabla o mediante su diagrama (que indica cómo puede componerse a partir de máquinas más sencillas). He aquí, por ejemplo, tanto la tabla como el diagrama de R y L:



En lo sucesivo, y siguiendo a Hermes, describiremos las máquinas de Turing sólo mediante diagramas, que resultan mucho más claros e intuitivos que las tablas cuando se trata de máquinas complicadas. Las máquinas que vamos a presentar son K (máquina de copiar la expresión anterior), ℛ (máquina de ir a la derecha de una sucesión de palabras; se para en el primer cuadro vacío del primer par de cuadros vacíos seguidos a la derecha del cuadro inicial), K_n (máquina de copiar la n-ava expresión, por la izquierda; por tanto $K_1 = K$), T (máquina de trasladar una expresión un cuadro a la izquierda), C (máquina de correr una palabra hasta el lugar que ahora ocupa otra palabra a su izquierda) y A (máquina de acabar, borrando los resultados intermedios). He aquí las descripciones esquemáticas de su acción y sus diagramas correspondientes (obsérvese que algunos signos de máquina llevan un exponente; éste indica el número de veces seguidas que ha de funcionar la máquina):



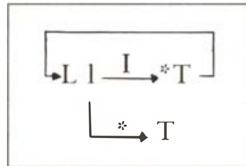
$$K_n : \frac{\ldots {}^*W_n{}^*W_{n-1}{}^*\ldots{}^*W_1\overset{*}{\uparrow}\ldots}{\ldots{}^*W_n{}^*W_{n-1}{}^*\ldots{}^*W_1{}^*W_n\overset{*}{\uparrow}}$$



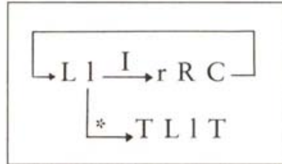
$$T : \frac{\tilde{\uparrow}\,{}^*W^*}{\sim W\overset{*}{\uparrow}\,{}^*}$$



$$C : \frac{{}^*W_2{}^*W_1\overset{*}{\uparrow}\ldots}{{}^*W_1\overset{*}{\uparrow}\,{}^*\ldots\;\ldots}$$



$$A : \frac{\sim{}^{**}W_1{}^*W_2{}^*\ldots{}^*W_n{}^*W_{n+1}\overset{*}{\uparrow}}{\sim W_{n+1}\overset{*}{\uparrow}\,{}^*\ldots\ldots\ldots\ldots\ldots\ldots\ldots}$$



Con esto tenemos aquí bastante. Designemos la fila de n + 1 palotes que representa el número natural n mediante «$\overline{n}$». Ahora nos encontramos en posición de dar la definición precisada de computabilidad recursiva en el sentido de Turing (o Turing-computabilidad) de una función numérica:

La función numérica f es Turing-computable si y sólo si hay una máquina de Turing M que cumple la siguiente condición: para cualquier número natural n, si escribimos $\overline{n}$ en la cinta de tal manera que todos los cuadros a su derecha estén vacíos y elegimos como primer cuadro de trabajo el primer cuadro vacío a la derecha de $\overline{n}$, M se pone a funcionar y, tras un número finito de pasos (en los que nunca va más hacia la izquierda que el primer cuadro vacío a la izquierda de $\overline{n}$), se para en el primer cuadro vacío a la derecha de f(n), estando f(n) separada de $\overline{n}$ por un cuadro vacío y encontrándose el resto derecho de la cinta vacío. (Lo mismo, *mutatis mutandis,* en el caso de varios argumentos). Es decir:

$$M : \frac{\ldots|{}^*\overline{n}\overset{*}{\uparrow}\ldots\ldots}{\ldots|{}^*\overline{n}{}^*\overline{f(n)}\overset{*}{\uparrow}\ldots\ldots}$$

La mayoría de las funciones numéricas familiares (adición, multiplicación, exponenciación, factorial, máximo, mínimo, etc.) pertenecen a una clase especialmente simple y rica a la vez de funciones computables: la clase de las funciones recursivas primitivas.

Las funciones recursivas primitivas son las funciones numéricas obtenibles a partir de las funciones recursivas primitivas iniciales (la función constante 0-ádica $C_0^0 = 0$, la función del siguiente, $S(x) = x' = x + 1$, y para cada dos números naturales $n (n \geqslant 1)$ e i $(1 \leqslant i \leqslant n)$, las funciones n-ádicas de identificación del i-avo miembro, $I_i^n(x_1,\ldots,x_n) = x_i$) mediante un número finito de aplicaciones de los procesos de definición por substitución y de definición por inducción.

Sea g una función r-ádica $(r \geqslant 1)$ y sean $h_1 \ldots h_r$ funciones n-ádicas $(n \geqslant 0)$. Decimos que la función n-ádica f está definida por substitución con ayuda de g, $h_1,\ldots,h_r$ si y sólo si para cualesquiera números naturales $x_1,\ldots,x_n$ ocurre que:

$$f(x_1,\ldots,x_n) = g(h_1(x_1,\ldots,x_n),\ldots,h_r(x_1,\ldots,x_n)).$$

Sea g una función n-ádica $(n \geqslant 0)$ y sea h una función $(n+2)$-ádica. Decimos que la función $(n+1)$-ádica f está definida por inducción con ayuda de g y h si y sólo si para cualesquiera números naturales $x_1,\ldots,x_n$, y ocurre que:

$$f(x_1,\ldots,x_n,0) = g(x_1,\ldots,x_n)$$
$$f(x_1,\ldots,x_n,y') = h(x_1,\ldots,x_n,y,f(x_1,\ldots,x_n,y)).$$

Hemos dicho que todas las funciones recursivas primitivas son computables (aunque no a la inversa). Si la precisión de Turing es adecuada, tendremos que poder mostrar que efectivamente todas esas funciones son Turing-computables. Es lo que hacemos a continuación, procediendo heurísticamente, para facilitar la comprensión. El siguiente apartado está, pues, dedicado a probar el teorema de que todas las funciones recursivas primitivas son Turing-computables.

Turing-computabilidad de las funciones recursivas primitivas

Las funciones recursivas primitivas son las funciones numéricas obtenibles a partir de las funciones recursivas primitivas iniciales mediante un número finito de aplicaciones de los procesos de definición por substitución y de definición por inducción. Por tanto, para probar que todas las funciones recursivas primitivas son Turing-computables hemos de probar que: 1. Las funciones recursivas primitivas iniciales son Turing-computables. 2. El proceso de definición por substitución lleva de funciones Turing-computables a funciones Turing-computables. 3. El proceso de definición por inducción conduce de funciones Turing-computables a funciones Turing-computables.

Lema 1: Las funciones recursivas primitivas iniciales son Turing-computables

Las funciones recursivas primitivas iniciales son la función constante $C_0^0 = 0$, la función del siguiente, $S(x) = x'$, y para cada dos números naturales $n (n \geqslant 1)$ e i $(1 \leqslant i \leqslant n)$, la función de identificación del i-avo miembro de la sucesión, $I_i^n (x_1,\ldots,x_n) = x_i$. Estas funciones son Turing-computables por las siguientes máquinas:

Función		máquina para computarla
C_0^0	:	r I r
$S(x) = x'$	:	K I r
$I_i^n (x_1,\ldots,x_n) = x_i$	:	K_{n+1-i}

Lema 2: El proceso de definición por substitución lleva de funciones Turing-computables a funciones Turing-computables

Empecemos por considerar un ejemplo. Sea g una función 3-ádica y sean h_1, h_2 y h_3 funciones 2-ádicas. Las funciones g, h_1, h_2 y h_3 sean Turing-calculables por las máquinas M_g, M_{h_1}, M_{h_2} y M_{h_3}, respectivamente. La función 2-ádica f esté definida por substitución con ayuda de g, h_1, h_2 y h_3 del siguiente modo:

$$f(x_1,x_2) = g(h_1(x_1,x_2),\ h_2(x_1,x_2),\ h_3(x_1,x_2))$$

Diseñemos una máquina de Turing para computar esta función f. Al comenzar, la situación de la cinta será

1) $*\bar{x}_1*\bar{x}_2\overset{*}{\uparrow}****\ldots$

es decir, al principio están los dos argumentos y el resto de la cinta está vacío. Primeramente construiremos un «puente» con un palote y copiaremos los argumentos

2) $*\bar{x}_1*\bar{x}_2*|*\bar{x}_1*\bar{x}_2\overset{*}{\uparrow}**\ldots$

A continuación borraremos el palote intermedio (creando así un intervalo de tres cuadros vacíos que señalará a la máquina de acabar, A, dónde habrá de pararse, cuando hayamos obtenido el resultado final), volviendo detrás de los argumentos copiados

3) $*\bar{x}_1*\bar{x}_2***\bar{x}_1*\bar{x}_2\overset{*}{\uparrow}\ldots$

Ahora computamos el valor de h_1 para x_1 y x_2, $h_1(x_1,x_2)$:

4) $*\bar{x}_1*\bar{x}_2***\bar{x}_1*\bar{x}_2*\overline{h_1(x_1,\ x_2)}\overset{*}{\uparrow}\ldots$

Volvemos a copiar los argumentos y computamos $h_2(x_1,x_2)$:

5) $*\bar{x}_1*\bar{x}_2***\bar{x}_1*\bar{x}_2*\overline{h_1(x_1,\ x_2)}*\bar{x}_1*\bar{x}_2*\overline{h_2(x_1,\ x_2)}\overset{*}{\uparrow}\ldots$

De nuevo copiamos los argumentos y computamos $h_3(x_1,x_2)$:

6) $*\overline{x}_1* \quad *\overline{h_1(x_1, x_2)}*\overline{x}_1*\overline{x}_2*\overline{h_2(x_1, x_2)}$
$* \quad *\overline{x}_2*\overline{h_3(x_1, x_2)}\overset{*}{\uparrow}$.

Ahora reunimos los valores $h_1(x_1, x_2)$, $h_2(x_1, x_2)$ y $h_3(x_1, x_2)$, a fin de poder computar el valor de la función g para ellos, tomados como argumentos:

7) $* \quad *\overline{x}_2** \quad *\overline{x}_2*\overline{h_1(x_1, x_2)}* \quad *\overline{x}_2*\overline{h_2(x_1, x_2)}$
$* \quad *\overline{x}_2*\overline{h_3(x_1, x_2)*h_1(x_1, x_2)*h_2(x_1, x_2)*h_3(x_1, x_2)}\overset{*}{\uparrow}*$

Con esto estamos en posición de computar $g(h_1(x_1, x_2), h_2(x_1, x_2), h_3(x_1, x_2))$, es decir, de $f(x_1, x_2)$:

8) $* \quad *\overline{x}_2** \quad *\overline{x}_2*\overline{h_1(x_1, x_2)}*\overline{x}_1*\overline{x}_2*\overline{h_2(x_1, x_2)}*\overline{x}_1*\overline{x}_2*$
$\overline{h_3(x_1, x_2)*h_1(x_1, x_2)*h_2(x_1, x_2)*h_3(x_1, x_2)*f(x_1, x_2)}\overset{*}{\uparrow}*$

Ahora sólo nos queda acabar, borrando los resultados intermedios y llevando el resultado final, $f(x_1, x_2)$, detrás de los argumentos.

9) $*$

¿Cómo diseñar una máquina para pasar de 1) a 7)? Claramente se ve que

el paso de 1) a 2) lo realiza	$r\|r\ K_3^2$
el paso de 2) a 3) lo realiza	$L^2\ l*ℜ$
el paso de 3) a 4) lo realiza	M_{h_1}
el paso de 4) a 5) lo realiza	$K_3^2\ M_{h_2}$
el paso de 5) a 6) lo realiza	$K_3^2\ M_{h_3}$
el paso de 6) a 7) lo realiza	$K_7\ K_5\ K_3$
el paso de 7) a 8) lo realiza	M_g
el paso de 8) a 9) lo realiza	A

Uniendo estas máquinas parciales obtenemos una máquina M_f para computar la función f:

$r|r\ K_3^2\ L^2\ l * ℜ M_{h_1}\ K_3^2\ M_{h_2}\ K_3^2\ M_{h_3}\ K_7\ K_5\ K_3\ M_g\ A.$

Ahora bien, el lema 2 no se limita al caso considerado en nuestro ejemplo en que f era una función 2-ádica definida por substitución con ayuda de 3 funciones 2-ádicas, sino que abarca todos los casos de definición por substitución.

Sea g una función r-ádica $(r \geqslant 1)$ y sean $h_1,\ldots,h_r$ funciones n-ádicas $(n \geqslant 0)$. Las funciones g, $h_1,\ldots,h_r$ sean Turing-computables por las máquinas M_g, $M_{h_1},\ldots,M_{h_r}$ respectivamente. La función n-ádica f esté definida por substitución con ayuda de g, $h_1,\ldots,h_r$ del siguiente modo:

$f(x_1,\ldots,x_n) = g(h_1(x_1,\ldots,x_n),\ldots,h_r(x_1,\ldots,x_n))$

Hemos de probar ahora que f es Turing-computable. Y, en efecto, la siguiente máquina M_f sirve para computar f:

$$r|r\ K^{n}_{n+1}\ L^{n}\ 1\ *\ (RM_{h_1}\ K^{n}_{n+1}\ M_{h_2}\ K^{n}_{n+1}\ \ldots\ M_{h_r}\ K_{r+(r-1)n}\ K_{r+(r-2)n}$$
$$\ldots K_r\ M_g\ A$$

como fácilmente se comprueba por consideraciones parecidas a las anteriormente expuestas. Obsérvese que, para $r=3$ y $n=2$, la máquina aquí indicada es idéntica a la obtenida en el ejemplo antes considerado.

Lema 3: El proceso de definición por inducción conduce de funciones Turing-computables a funciones Turing-computables

Empecemos también aquí por considerar un ejemplo. Sea g una función 1-ádica y sea h una función 3-ádica. Las funciones g y h sean Turing-computables por las máquinas M_g y M_h, respectivamente. La función 2-ádica f esté definida por inducción con ayuda de g y h del siguiente modo:

$f(x,\ 0)=g(x)$
$f(x,y')=h(x,y,f(x,y))$

Diseñemos una máquina de Turing para computar esta función f. Como siempre, al comenzar, los dos argumentos estarán al principio de la cinta; el resto estará vacío:

1) $*\overline{x}*\overline{y}\overset{*}{\uparrow}**\ldots$

Empecemos por construir un «puente» con un palote y copiemos los argumentos en orden inverso:

2) $*\overline{x}*\overline{y}*|*\overline{y}*\overline{x}\overset{*}{\uparrow}\ldots$

Borremos ahora el palote intermedio (creando así un intervalo de tres cuadros vacíos que señalará a la máquina de acabar, A, dónde habrá de pararse, cuando hayamos obtenido el resultado final) y volvamos detrás de los argumentos copiados:

3) $*\overline{x}*\overline{y}***\overline{y}*\overline{x}\overset{*}{\uparrow}\ldots$

¿Cómo proceder? Iremos computando sucesivamente f(x,0), f(x,1), f(x,2), f(x,3), etc. hasta llegar a f(x,y), que será nuestro resultado final. Comencemos con f(x,0), que, como sabemos, es igual a g(x). Bastará, pues, con computar g(x):

4) $*\overline{x}*\overline{y}***\overline{y}*\overline{x}*\overline{f(x,\ 0)}\overset{*}{\uparrow}\ldots$

Si $y=0$, ya hemos obtenido el resultado final. Si $y\neq 0$, deberemos proseguir. ¿Cómo saber si $y=0$ o no? Copiando $\overline{y}$ y quitándole un palote. Si no queda ninguno, es que y era 0 (representado por un solo palote). Si aún quedan palotes, es que y era distinto de 0.

Supongamos que $y \neq 0$. Ahora debemos computar f(x,1) = h(x,0,f(x,0)). Para ello copiaremos x, escribiremos 0 y copiaremos f(x,0), computando a continuación el valor de la función h, aplicada a x, 0 y f(x,0), con lo que tendremos f(x,1):

5) $*\overline{x}*\overline{y}***\overline{y}*\overline{x}*\overline{f(x,0)}*\overline{y-1}*\overline{x}*\overline{0}*\overline{f(x,0)}*\overline{f(x,1)}\overset{*}{\uparrow}\ldots$

Si $y=1$, ya hemos obtenido el resultado final. Si $y>1$, deberemos proseguir. ¿Cómo saber si $y=1$ o no? Copiando $\overline{y-1}$ (que es la quinta palabra, contando de derecha a izquierda) y quitándole un palote.

Si no queda ninguno, es que $y-1$ era 0 (representado por un solo palote) y, por tanto, que y era 1. Si aún quedan palotes, es que y era mayor que 1.

Supongamos que $y \neq 1$. Ahora debemos computar f(x,2) = h(x,1,f(x,1)). Para ello copiaremos $\overline{x}$, escribiremos $\overline{1}$ (para lo que bastará copiar $\overline{0}$, que en ese momento será la quinta palabra, contando de derecha a izquierda, y añadirle un palote) y copiaremos $\overline{f(x,1)}$, computando a continuación el valor de la función h, aplicada a x, 1 y f(x,1), con lo que tendremos f(x,2):

6) $*\overline{x}*\overline{y}***\overline{y}*\overline{x}*\overline{f(x,0)}*\overline{y-1}*\overline{x}*\overline{0}*\overline{f(x,0)}*\overline{f(x,1)}*\overline{y-2}*\overline{x}*\overline{1}*$
$\overline{f(x,1)}*\overline{f(x,2)}\overset{*}{\uparrow}**\ldots$

Si $y=2$, ya hemos obtenido el resultado final. Si $y>2$, deberemos proseguir. ¿Cómo saber si $y=2$ o no? Copiando $\overline{y-2}$ (que es la quinta palabra, contando de derecha a izquierda) y quitándole un palote. Si no queda ninguno, es que $y-2$ era 0 (representado por un solo palote) y, por tanto, que y era 2. Si aún quedan palotes, es que y era mayor que 2.

Supongamos que $y \neq 2$. Ahora debemos computar f(x,3)= h(x,2,f(x,2)). Para ello copiaremos $\overline{x}$, escribiremos $\overline{2}$ (para lo que bastara copiar $\overline{1}$, que en ese momento será la quinta palabra, contando de derecha a izquierda, y añadirle un palote) y copiaremos $\overline{f(x,2)}$, computando a continuación el valor de la función h, aplicada a x, 2 y f(x,2), con lo que tendremos f(x,3):

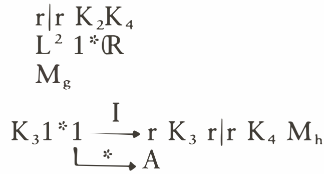
7) $*\overline{x}*\overline{y}***\overline{y}*\overline{x}*\overline{f(x,0)}*\overline{y-1}*\overline{x}*\overline{0}*\overline{f(x,0)}*\overline{f(x,1)}*\overline{y-2}*\overline{x}*\overline{1}*$
$\overline{f(x,1)}*\overline{f(x,2)}*\overline{y-3}*\overline{x}*\overline{2}*\overline{f(x,2)}*\overline{f(x,3)}\overset{*}{\uparrow}\ldots$

Si $y=3$, ya hemos obtenido el resultado final. Si $y>3$, deberemos proseguir. Está claro que seguiremos el mismo proceso para computar f(x,4), f(x,5),... hasta llegar a f(x,y). Entonces acabaremos, borrando las anotaciones y resultados intermedios y llevando el resultado final hasta el intervalo de tres cuadros vacíos construidos al principio, de modo que finalmente obtengamos

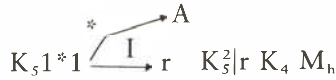
(fin) $*\overline{x}*\overline{y}*f(x,y)\overset{*}{\uparrow}**\ldots$

¿Cómo diseñar una máquina de Turing para pasar de (1) a (fin)? Claramente se ve que

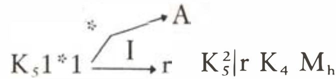
el paso de (1) a (2) lo realiza
el paso de (2) a (3) lo realiza
el paso de (3) a (4) lo realiza
el paso de (4) a (5) (o a (fin)) lo realiza



el paso de 5) a 6) (o a (fin)) lo realiza

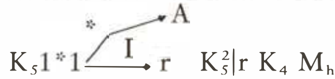


el paso de 6) a 7) (o a (fin)) lo realiza

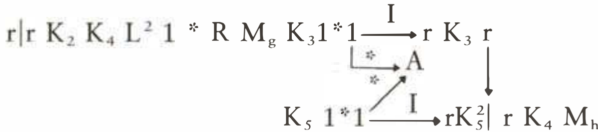


....
....
....

el paso de (n) a (n + 1) (o a (fin)) lo realiza



Vemos que, a partir del paso de 5) a 6), la misma máquina parcial realiza todos los pasos a dar. Por tanto, uniendo las primeras máquinas parciales y empalmando la cuarta con la quinta (que es la que se repite hasta el final), obtenemos una máquina M_f para computar la función f:

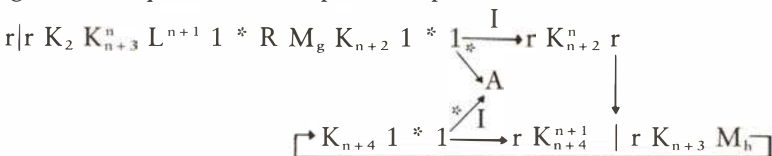


De todos modos, el Lema 3 no se limita al caso considerado en nuestro ejemplo en que f era una función 2-ádica definida por inducción con ayuda de una función l-ádica y otra 3-ádica, sino que abarca todos los casos de definición por inducción.

Sea g una función n-ádica ($n \geqslant 0$) y sea h una función $(n+2)$-ádica. Las funciones g y h sean Turing-computables por las máquinas M_g y M_h, respectivamente. La función $(n+1)$-ádica f está definida por inducción con ayuda de g y h del siguiente modo:

$f(x_1,\ldots,x_n,\ 0) = g(x_1,\ldots,x_n)$
$f(x_1,\ldots,x_n,\ y') = h(x_1,\ldots,x_n,\ y,\ f(x_1,\ldots,x_n,\ y))$

Hemos de probar ahora que f es Turing-computable. Y, en efecto, la siguiente máquina M_f sirve para computar f:



como fácilmente se comprueba por consideraciones parecidas a las anteriormente expuestas. Obsérvese que, para $n=1$, la máquina aquí indicada es idéntica a la obtenida en el ejemplo que acabamos de considerar.

De los Lemas 1, 2 y 3 y de la definición de función recursiva primitiva claramente se sigue lo que queríamos probar, a saber:

Teorema: Todas las funciones primitivas recursivas son Turing-computables.

V. FUNCIONES RECURSIVAS

Aunque todas las funciones recursivas primitivas son computables, no todas las funciones computables son recursivas primitivas. Por ejemplo, la siguiente función f, definida por Ackermann en 1928, es computable en sentido intuitivo (y Turing-computable), pero no recursiva primitiva (recuérdese que s es la función del sucesor):

$f(0, y) = s(y)$
$f(s(x), 0) = f(x, 1)$
$f(s(x), s(y)) = f(x, f(s(x), y))$

Una noción más amplia es la de función recursiva, cuya definición requiere la previa introducción del operador *min* (el mínimo... tal que). Hablando de números naturales, *min* x $\varphi(x)$ es el mínimo número x que satisface la condición φ. Si hay algún número que satisface φ, y para cada número natural x es decidible si $\varphi(x)$ o no, entonces *min* x $\varphi(x)$ es computable. Decimos que una función n-aria h es definible por minimalización a partir de una función $n+1$-aria f en caso normal si y sólo si para cada $x_1,\ldots,x_n$ existe al menos un w tal que $f(x_1,\ldots,x_n, w)=0$, y ocurre que para cada $x_1,\ldots,x_n$:

$h(x_1,\ldots,x_n) = \mathit{min}\ w\,[f(x_1,\ldots,x_n, w)=0]$

Una función recursiva es una función definible a partir de las funciones recursivas primitivas iniciales por un número finito de definiciones por sustitución, por inducción y por minimalización en caso normal.

La función de Ackermann es recursiva. De hecho toda función computable conocida es recursiva. Y se puede probar el siguiente teorema: Una función es recursiva si y sólo si es Turing-computable.

El hecho de que las nociones de recursividad y computabilidad de Turing coincidan (aun partiendo de ideas iniciales bien distintas), de que todas las otras precisiones de la noción de computabilidad propuestas por otros autores (como Post, Markov y Church) hayan resultado también equivalentes, y de que toda función conocida y computable en sentido intuitivo sea también Turing-computable, ha llevado a la conclusión (conocida como tesis de Church) de que el concepto intuitivo de computabilidad queda perfectamente precisado por la noción de computabilidad de Turing (o de recursividad).

VI. ALGUNOS RESULTADOS SOBRE LA LÓGICA

La teoría de la computabilidad sirve para establecer resultados sobre decidibilidad y generabilidad (o numerabilidad recursiva) de diversas partes de la lógica. Aquí nos limitaremos a mencionar algunos, sin prueba ni comentario:

La lógica conectiva o proposicional es decidible (es decir, el conjunto de sus fórmulas válidas es decidible respecto al conjunto de todas sus fórmulas).

El fragmento de la lógica de primer orden que sólo usa predicados monarios (y que incluye la silogística) es decidible.

La lógica de primer orden es indecidible.

La lógica de segundo orden es indecidible.

Cualquier lógica decidible (como la conectiva o la de predicados monarios de primer orden) es recursivamente numerable.

La lógica de primer orden (aunque no decidible) es recursivamente numerable. Por eso hay cálculos deductivos que generan todas sus fórmulas válidas.

La lógica de segundo orden no es recursivamente numerable. Por eso no puede haber cálculos deductivos adecuados para ella.

Un conjunto de fórmulas de la lógica de primer orden constituye una clase de reducción si y sólo si la decidibilidad de ese conjunto implicaría la de la lógica de primer orden. Por tanto, todas las clases de reducción son indecidibles. Por ejemplo, Kalmar probó en 1936 que el conjunto de todas las fórmulas válidas con un solo predicado binario es una clase de reducción y, por tanto, indecidible. Desde entonces el problema de la decisión de la lógica de primer orden ha sido exhaustivamente analizado a base de tipificar los subconjuntos de fórmulas, clasificados por su prefijo en forma normal, y descubrir cuáles son decidibles y cuáles son clases de reducción, indecidibles.

VII. ALGUNOS RESULTADOS SOBRE TEORÍAS

Una teoría es un conjunto de fórmulas clausurado respecto a la relación de consecuencia lógica, es decir, un conjunto de fórmulas que incluye todas sus consecuencias.

Una teoría se llama axiomatizable si y sólo si es recursivamente numerable.

Una teoría se llama completa si y sólo si da respuesta a todas las preguntas que se pueden formular en su lenguaje, es decir, si para cada fórmula φ de su lenguaje, $\varphi \in T$ o $\neg \varphi \in T$.

Una teoría se llama consistente si y sólo si no incluye todas las fórmulas de su lenguaje.

Claramente, la axiomatizabilidad, la completud y la consistencia son propiedades deseables de una teoría. En 1931 probó Gödel su famoso

teorema de incompletud, que (en una versión generalizada) puede formularse así:

Una teoría matemática interesante (es decir, una en la que al menos sean definibles las funciones recursivas primitivas) no puede ser a la vez axiomatizable, completa y consistente. Puede ser completa y consistente, pero no axiomatizable (como la aritmética intuitiva); puede ser axiomatizable y completa, pero no consistente (como cualquier aritmética contradictoria); y, finalmente, puede ser axiomatizable y consistente, pero no completa (como la aritmética de Peano de primer orden). Este teorema nos dice que ciertos ideales son inalcanzables conjuntamente.

También se han obtenido numerosos resultados de decidibilidad e indecidibilidad acerca de teorías (respecto a sus lenguajes). He aquí algunos ejemplos, relativos a diversas teorías de primer orden:

La teoría pura de la igualdad es decidible.
La teoría de las álgebras de Boole es decidible.
La teoría de grupos abelianos es decidible.
La teoría de grupos abelianos ordenados es decidible.
La geometría hiperbólica es decidible.
La teoría de grupos es indecidible.
La teoría de retículos es indecidible.
La teoría de retículos distributivos es indecidible.
La teoría de cuerpos ordenados es indecidible.
La teoría de conjuntos (de Zermelo-Fraenkel) es indecidible.

BIBLIOGRAFÍA

Boolos, G. y Jeffrey, R. (1980), *Computability and Logic,* Cambridge University Press.

Church, A. (1936), «An unsolvable problem of elementary number theory»: *American Journal of Mathematics, 58.*

Cutland, N. (1980), *Computability. An introduction to recursive function theory,* Cambridge University Press.

Dreben, B. y Goldfarb, W. (1980), *Decision problems. Solvable Classes of Quantificational Formulas,* Addison-Wesley.

Gödel, K. (1931), «Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme»: *Monatshefte für Mathematik und Physik, 38* (v.e. K. Gödel, *Obras completas,* Alianza, [2]1989).

Hermes, H. ([2]1971), *Aufzählbarkeit, Entscheidbarkeit, Berechenbarkeit,* Springer-Verlag (v.e. *Introducción a la teoría de la computabilidad,* Tecnos, 1984).

Kalmar, L. (1936), «Zurückführung des Entscheidungsproblems auf den Fall von Formeln mit einer einzigen binären Funktionsvariablen»: *Compositio mathematica, 4.*

Monk. D. (1976), *Mathematical Logic,* Springer-Verlag.

Rogers, H., Jr. (1967), *Theory of Recursive Functions and Effective Computability,* McGraw-Hill Co.

Turing, A. (1937), «On computable numbers, with an application to the Entscheidungsproblem»: *Proceedings of the London Mathematical Society, 42-43.*

LÓGICA MODAL

Raúl Orayen

En las últimas décadas la investigación sobre lógica modal ha aumentado y se ha diversificado de tal manera, que se ha vuelto difícil dar un panorama de esta disciplina (para tener una idea de cómo cambió la situación en sólo dieciséis años, compárense los prefacios de Hughes y Cresswell, 1968 y Hughes y Cresswell, 1984; véase también Bull y Segerberg, 1984, 2). En este artículo evitaré dos políticas extremas, la de dar un panorama de todo sin detalles, y la de estudiar unos pocos temas de manera minuciosa. Intentaré dar una visión general histórico-sistemática de los aspectos más importantes de la lógica modal, pero agregaré a la descripción panorámica un análisis algo más detallado de algunos sistemas y resultados formales de especial interés, o representativos de alguna línea de investigación mencionada en el texto. El objetivo del trabajo es doble: por un lado, dar un perfil de la lógica modal; por otro, suministrar una introducción a este campo a quienes deseen adquirir un conocimiento más profundo del mismo. Sólo presupondré conocimientos de la lógica elemental como los que proporciona Quesada (1991).

El plan del capítulo es el siguiente. La sección I se ocupa de preliminares conceptuales y en la II se hace un bosquejo del desarrollo histórico de la lógica modal. Siguen tres secciones sobre lógica modal proposicional (aunque de carácter sistemático, corresponden a tres etapas históricas mencionadas antes). En un apéndice se mencionan líneas de investigación que no son tratadas en el cuerpo principal del artículo. La bibliografía ha sido cuidadosamente seleccionada para facilitar investigaciones o estudios ulteriores. Se citan especialmente dos tipos de obras: trabajos clásicos en los cuales se expusieron por primera vez ideas o resultados importantes conectados con la lógica modal, y literatura expositiva valiosa por la presentación clara y sistemática de desarrollos que están dispersos en artículos técnicos. Todas las obras de la bibliografía se han mencionado en el texto y los comentarios que hago sobre ellas ayudarán

al lector interesado en alguno de los temas tratados a elaborar su propio plan de lecturas adicionales.

I. PRELIMINARES CONCEPTUALES

En la lógica medieval se pensaba que había diferentes *modos* en que una proposición podía ser verdadera o falsa. Considérense las proposiciones expresadas por las oraciones 'Sócrates conversaba mucho de filosofía' y 'Sócrates se interesaba por la filosofía o no es cierto que Sócrates se interesaba por la filosofía'. Ambas proposiciones son verdaderas, pero se pensaba que la primera lo es de un modo *contingente* y la segunda de un modo *necesario.* Similarmente, una proposición falsa puede serlo de diversos modos. La proposición expresada por 'Sócrates murió por ingerir estricnina' es falsa pero *posible,* en tanto que la expresada por '2 + 2 = 5' es falsa y además *imposible* (si los signos tienen sus significados usuales). En la terminología lógica actual (que en este punto conserva reminiscencias medievales) se dice que la *necesidad,* la *posibilidad,* la *contingencia* y la *imposibilidad* son *modalidades.* La rama de la lógica que se ocupa de ellas es la *lógica modal.* En el simbolismo lógico, las modalidades suelen representarse mediante operadores proposicionales. Veremos ahora los operadores más usados en lógica modal.

Introduciremos cuatro operadores proposicionales monádicos $\Box$, $\Diamond$, *C, I,* que corresponden a las modalidades antes mencionadas, se llaman por eso 'operadores modales', y pueden combinarse con las variables proposicionales *p, q, r,* eventualmente usadas con subíndices[1]. Tales combinaciones se leerán así:

$\Box p$:*p* es necesaria,
$\Diamond p$:*p* es posible,
Cp:p es contingente,
Ip:p es imposible.

En esta interpretación intuitiva del simbolismo introducido, los valores de la variable *p* son proposiciones y $\Box p$ es verdadero si y sólo si *p* es una proposición necesaria (similarmente para los otros operadores). Antes de comenzar con temas más formales, haremos algunas reflexiones sobre el significado de los operadores introducidos y mencionaremos algunas discusiones filosóficas a que han dado lugar.

Si se entiende el operador $\Box$ de acuerdo con la explicación del párrafo anterior, su significado será claro sólo si resulta clara la noción de *proposición necesaria.* Durante este siglo, los filósofos analíticos trataron de aclarar esa noción (y en general, las nociones modales básicas). Surgieron entonces dos complicaciones en la filosofía de la lógica modal. En primer lugar, resultó difícil explicar el significado de 'proposición nece-

1. Los signos y expresiones lógicas serán usados a menudo como nombres de sí mismos, aunque cuando parezca conveniente usaré también comillas simples del modo usual en lógica.

saria' y algunos autores adoptaron una suerte de *escepticismo intensional-modal* que niega la existencia de proposiciones y la significatividad de la noción de *necesidad*. En segundo lugar, los lógicos y filósofos que sí aceptaron la noción de *necesidad* no dieron una sola explicación de ella: durante este siglo ha habido al menos dos posiciones muy diferentes acerca de cómo debía interpretarse ese concepto en el contexto teórico de la lógica modal (las dos interpretaciones corresponden aproximadamente a la distinción tradicional entre modalidades *de dicto* y *de re*). Antes de ver estas dos cuestiones filosóficas, será conveniente hacer notar algunas relaciones entre los operadores modales que simplifican un poco los problemas conceptuales conectados con ellos.

Hemos dicho que no ha resultado fácil aclarar el significado de las nociones modales. Pero un hecho muy importante es que resultan mucho más claras las relaciones entre ellas. Tomando como primitiva una cualquiera de las modalidades antes mencionadas, se pueden definir todas las demás en términos de la escogida. Con respecto a los operadores modales antes introducidos puede decirse entonces lo siguiente: si se toma uno cualquiera de los cuatro, los restantes se pueden definir en términos del operador elegido y los conectivos veritativo-funcionales de la lógica proposicional. Es habitual tomar como base el operador de necesidad. En ese caso, los otros operadores pueden definirse así:

(*D*1) $\Diamond p =_{\text{def.}} \sim \Box \sim p$

(*D*2) $Ip =_{\text{def.}} \Box \sim p$

(*D*3) $Cp =_{\text{def.}} \sim \Box p \wedge \sim \Box \sim p$

(Puede resultar un ejercicio divertido para el lector ir eligiendo como básico cada uno de los otros operadores y en cada caso definir los restantes en términos del elegido. Es interesante la definición de $\Box$ usando *C*).

Los operadores que hemos introducido hasta ahora son *monádicos;* hay también operadores modales *binarios,* de los que nos ocuparemos en futuras secciones. El hecho de que los operadores modales introducidos sean interdefinibles (y de que haya mucho consenso sobre la adecuación de las definiciones antes formuladas) simplifica enormemente la discusión de las dos «complicaciones» filosóficas mencionadas unos párrafos más atrás (además del interés intrínseco de tal interdefinibilidad). No hay que discutir con el escéptico «intensional-modal» operador por operador: si se logra convencerlo de que uno de los operadores monádicos tiene una significatividad clara, aceptará que también los demás la tienen, porque se pueden introducir mediante definiciones sencillas a partir del operador aceptado. Similarmente, si se discute cuál de dos concepciones de la necesidad es adecuada para interpretar cierta teoría modal, ya no será necesario discutir lo mismo respecto de las otras modalidades monádicas: debido a las íntimas relaciones que se dan entre ellas, diferentes concepciones de la necesidad serán paralelas a distintas concepciones de la contingencia, etc., y la elección de una concepción de una de las modalidades llevará naturalmente a elegir la concepción afín de otra de ellas.

Debido a estas razones, al sintetizar en lo que sigue algunas discusiones filosóficas, nos ocuparemos de una sola modalidad monádica (o de un solo operador monádico). Tomaremos como básico el concepto de *necesidad* —o el operador □. Como hemos dicho, hay filósofos que han rechazado esta noción y otros han dado explicaciones divergentes de ella. Ampliaremos esta información.

En la interpretación de la lógica modal, ha habido dos concepciones influyentes de la *necesidad.* La primera de ellas fue desarrollada por Carnap (1947) y refinada en Carnap (1956a). El núcleo de este enfoque es la idea de que una proposición es necesaria si cualquier oración que la exprese es analítica, y una oración es analítica si las reglas semánticas bastan para establecer su verdad. Si el operador de necesidad ('N', en la notación de Carnap) se prefija a una oración, el resultado es verdadero si y sólo si la oración es analítica. Esta explicación encapsula una concepción de la *necesidad* que tuvo enorme influencia. Al formularla, Carnap (1956a, 174) utiliza la expresión 'L-verdadera' en lugar de 'analítica'; pero en su libro, 'L-verdadera' se toma en un sentido amplio: se aplica a lo que hoy llamamos 'lógicamente verdadero', pero también a oraciones como 'ningún soltero es casado'. En otras palabras, 'L-verdadera' se usa como 'analítica' (incluyendo lo lógicamente verdadero como un caso particular). De acuerdo con este enfoque, 'N(ningún soltero es casado)' es verdadera, ya que 'ningún soltero es casado' es analítica, y esto último se cumple porque bastan las reglas semánticas para establecer la verdad de esa oración. Llamaremos 'concepción semántica de la necesidad' a la propuesta por Carnap, en vista de que se explica en términos de propiedades semánticas de las oraciones. Es muy importante advertir que los lógicos actuales usan una noción de *necesidad* que es esencialmente idéntica a la de Carnap: es la llamada 'necesidad lógica'. Se dice en los textos lógicos que un razonamiento *R* es *válido* cuando cumple con esta condición: si las premisas de *R* son verdaderas, entonces necesariamente la conclusión de *R* es verdadera. 'Necesariamente' se usa aquí en el sentido de la *necesidad lógica.* Otra definición usual dice que un razonamiento es válido cuando no es *posible* que sus premisas sean verdaderas y su conclusión falsa. También esta noción de posibilidad se usa en el sentido lógico y se la define en términos de la *necesidad lógica* mediante (*D*1) (definiendo así la posibilidad lógica, las dos definiciones de *validez* ofrecidas son equivalentes en los contextos lógicos usuales; véase Orayen, 1989, sección 4.1).

La concepción de la necesidad propuesta por Carnap fue objeto de numerosas críticas a lo largo de los años. Quine, adalid del escepticismo intensional-modal mencionado antes, le hizo dos objeciones clásicas. Una de ellas, de carácter muy general, se basa en el conocido rechazo quineano de la distinción analítico-sintético, formulado en el célebre Quine (1951), y una de cuyas consecuencias es el rechazo de todo concepto definido sobre la base de la noción de *analiticidad* (como la *necesidad* carnapiana). Como mencionamos antes, Carnap (1956a) es una versión mejo-

rada de Carnap (1947), y uno de los agregados que contiene es Carnap (1956b), que constituye un intento de responder al último artículo mencionado de Quine. Quine no se dejó convencer y profundizó sus críticas contra el concepto de *analiticidad* (y nociones «intensionales»[2] relacionadas) mediante el llamado «argumento de la indeterminación de la traducción», expuesto en Quine (1960, C.2) y numerosos trabajos posteriores. No hay consenso acerca del peso que se les deba dar a estas objeciones anti-intensionalistas. En Orayen (1989) puede hallarse un análisis crítico detenido de los argumentos de Quine contra los llamados «conceptos intensionales» (véanse los capítulos 2 y 3, y el Apéndice I, que es una respuesta de Quine a una versión previa del capítulo 2).

Otra célebre objeción de Quine a la noción de *necesidad* carnapiana no depende en absoluto de la comentada en el párrafo anterior —de hecho, fue formulada cuando Quine no había roto aún con la noción de *analiticidad.* En efecto, en su trabajo (1943), Quine había señalado que si en la lógica modal se combinaba una cuantificación objetal (la usual en la lógica clásica de primer orden) con una noción de *necesidad* basada en la de *analiticidad* (como la carnapiana), se obtenían confusiones inextricables. Por ejemplo, si se interpreta de manera objetal el cuantificador de

$(\exists x)\Box(x>7)$,

la fórmula resultará verdadera si algún objeto del dominio de variabilidad (supongamos que es el conjunto de los números naturales) satisface el alcance $\Box(x>7)$. Pero si $\Box$ se entiende de la manera carnapiana, no puede decirse que un objeto (el número 9, por ejemplo) satisface $\Box(x>7)$, porque eso parece depender de la expresión lingüística con la que nos refiramos al objeto. En efecto, '$\Box(9>7)$' parece «carnapianamente» verdadero porque '$(9>7)$' puede considerarse analítico, pero '$\Box$ (el número de planetas >7)' debe considerarse carnapianamente falso, porque la verdad de '(el número de planetas >7)' no depende sólo del significado de las palabras. Sin embargo, '9' y 'el número de planetas' denotan el mismo objeto (el número 9). ¿Ese objeto satisface o no $\Box(x>7)$? Parece que la respuesta es afirmativa si nos referimos al objeto mediante '9' y falsa si nos referimos a ese mismo objeto mediante 'el número de planetas'. Pero si la respuesta depende de la manera en que nos refiramos al objeto, parece que el objeto mismo no satisface ni deja de satisfacer $\Box(x>7)$. Si *el objeto mismo* (el número 9) satisficiera $\Box(x>7)$, se podría obtener una verdad de esa fórmula reemplazando la 'x' que figura en ella por *cualquier* término singular que denotara el número 9. Pero si un objeto del dominio de variabilidad se comporta en esa forma indecisa respecto de oraciones abiertas o funciones proposicionales del lenguaje modal carnapiano, hay confusiones semánticas serias en el tratamiento «carnapiano» de la lógica modal con cuantificadores.

2. Se trata de nociones estrechamente relacionadas con el *significado* (denominado 'intensión' en la terminología de Carnap y otros autores; el término se escribe con 's' para diferenciarlo del término psicológico similar).

Antes de ver otra concepción de la *necesidad,* será conveniente hacer algunas precisiones sobre las dos críticas de Quine que hemos considerado. Su primera objeción no está dirigida específicamente contra una noción modal usada por Carnap: es muy general y pone en tela de juicio *cualquier* noción basada en la de *analiticidad* (la *necesidad* carnapiana es objetada como un caso particular). Esta crítica de Quine tiene tanta fuerza como sus argumentos anti-intensionalistas, respecto de los cuales difieren las opiniones. La segunda objeción es mucho más específica y contundente: afecta a un tipo de lógica modal en que se mezcla la *cuantificación objetal* con una noción de *necesidad* basada en la *analiticidad,* y Quine señala correctamente que se comete entonces una seria confusión. Pero esto no es fatal para el desarrollo de la lógica modal. Hay muchas formas en que una teoría modal puede eludir este problema: dos caminos posibles son usar otro tipo de cuantificación u otro tipo de *necesidad*[3]. Quine no estaba interesado en sugerir el segundo camino: sus críticas propiciaban el abandono de la *necesidad,* más bien que el desarrollo de una concepción alternativa de esa noción. Pero en 1972 se publicó un trabajo de Kripke (ahora disponible como Kripke, 1980) que cambió totalmente el panorama teórico en la filosofía de la lógica modal, porque en él se critica la vinculación entre *necesidad* y *analiticidad* que se había hecho hasta ese momento, pero se propone también otra concepción de la *necesidad.*

La idea básica presentada en Kripke (1980) es que hay un tipo de necesidad metafísica que no se reduce de ningún modo a una necesidad semántica: un estado de cosas puede ser necesario aunque una oración que lo exprese no sea analítica. Es una verdad necesaria que Héspero es idéntico a Fósforo, pero éste es un hecho acerca de objetos: 'Héspero = Fósforo' no es analítica (los dos nombres propios allí usados no son sinónimos). Kripke apoya sus ideas en un minucioso análisis de ejemplos que resultaría difícil resumir. El meollo de uno de sus argumentos puede presentarse de la manera siguiente. Kripke trata de hacer ver al lector que tiene un sentido intuitivo preguntarse si un determinado objeto *x podría no haber tenido* una propiedad *P* que *de hecho tiene;* aduce además que una pregunta de este tipo no se ocupa de palabras ni de significados. Pero si la pregunta tiene sentido y no es de tipo semántico, parece que una respuesta negativa también será significativa y tampoco tendrá un carácter semántico. En ese caso tiene un sentido intuitivo una noción de necesidad ontológica, porque responder que un objeto *no* podría haber carecido de cierta propiedad es atribuírsela *necesariamente.* Kripke defiende la idea de que un objeto puede tener *necesariamente* algunas de

3. Durante un tiempo, Quine consideró correcta una tercera línea de solución, propuesta por Church (1943) y Carnap (1947), y basada en una limitación de los dominios de variabilidad a valores intensionales. Sin embargo, en la segunda edición de su *From a Logical Point of View* (1961a) señaló nuevas dificultades de estas propuestas. Quine (1961b) sintetiza estupendamente esta discusión triangular (en Quine, 1980, se atenúa la crítica a Church de manera no decisiva).

sus propiedades y afirma que éste no es un hecho lingüístico. Defiende entonces la significatividad de las oraciones en las que se afirma lo que en terminología tradicional se llamaba 'necesidad *de re*'. Esta posición se conoce ahora con el nombre de 'esencialismo' y contrasta fuertemente con la filosofía modal prevaleciente desde el positivismo lógico hasta Kripke. Tal filosofía modal trataba de reducir toda *necesidad* a la *de dicto:* la noción de *necesidad* se aplicaba a proposiciones, o mejor aun, a oraciones que las expresaran. El trabajo de Kripke antes mencionado tuvo una influencia notable, y a partir de su difusión la filosofía analítica se mostró más suspicaz acerca de la asimilación de toda necesidad a la *de dicto.* Como consecuencia de la mayor aceptación de la *necesidad de re,* resurgió en las dos últimas décadas el interés por los problemas metafísicos vinculados con las modalidades. Una obra interesante para ver estos problemas es Forbes (1985), donde el autor desarrolla una teoría propia acerca de la *necesidad de re,* pero proporcionando al lector la información básica que se requiere para poder seguir las discusiones filosóficas recientes sobre esa temática.

El lector puede preguntarse cuál fue la actitud de Quine ante el esencialismo. La situación es algo paradójica, porque Quine también rechazó esta concepción de la *necesidad* (aunque con argumentos menos contundentes: véase Sainsbury 1991, 242-243) pero aparentemente contribuyó a su desarrollo: como señala Kaplan en un penetrante trabajo sobre el tratamiento quineano de la cuantificación en contextos opacos, Quine sugirió claramente que para dar sentido a fórmulas modales como la que antes analizamos, había que interpretarlas de manera esencialista (véase Kaplan, 1968). La necesidad de tal interpretación era para Quine una prueba de que tales fórmulas eran defectuosas, pero curiosamente otros autores pudieron extraer de la crítica quineana una sugerencia positiva.

Las dos concepciones de la *necesidad* que he explicado aquí (la necesidad semántica o lógica y la necesidad metafísica) no son las únicas que se han propuesto (ni los autores mencionados son los únicos que las han defendido), pero son las que han tenido más repercusión en la historia reciente de la lógica modal y las especulaciones filosóficas sobre esa disciplina.

II. ESBOZO HISTÓRICO DE LA LÓGICA MODAL

Dividiré la historia de la lógica modal en cuatro períodos: la prehistoria, la etapa sintáctica, la etapa semántica y la época de la metalógica modal generalizada. La prehistoria abarca desde el tiempo de Aristóteles (384-322 a. C.) hasta 1912, año en que C. I. Lewis inauguró la historia moderna de la disciplina. La etapa sintáctica, caracterizada por el surgimiento de sistemas axiomáticos modales que, en general, eran presentados sin una semántica sistemática, se extiende desde 1912 hasta 1959. Ese año Kripke comienza a publicar trabajos sobre la semántica de mun-

dos posibles y comienza la etapa semántica, durante la cual se investiga la aplicación de los métodos de Kripke a varios sistemas particulares que se habían estudiado de manera sintáctica en el período anterior. La etapa en la que pienso al hablar de «la época de la metalógica modal generalizada» no tiene un comienzo muy nítido en el tiempo, pero lo podemos situar hacia fines de la década del sesenta. El rasgo que enfatizo con el rótulo elegido es la generalidad. En esta época, hasta la actualidad, el interés por sistemas particulares es reemplazado en gran parte por el intento de estudiar las propiedades de clases muy amplias de sistemas modales. Daré alguna breve información de lo que ocurrió en las distintas etapas mencionadas.

1. *La prehistoria de la lógica modal*

La palabra 'prehistoria' del subtítulo se usa sin intenciones peyorativas. Simplemente intento enfatizar con ella el hecho de que en este largo período la reflexión sobre las modalidades no arrojó como resultado ningún sistema axiomático descripto con claridad, y con una interpretación intuitiva de tipo modal. En la actualidad, estos rasgos son las condiciones *mínimas* que una teoría formal debería reunir para ser considerada un sistema de lógica modal (obsérvese la modestia de la exigencia semántica). Es natural, entonces, que desde nuestro punto de vista contemporáneo, consideremos que, estrictamente hablando, la historia de la lógica modal en un sentido moderno no había comenzado aún en un período sin logros del tipo mencionado dos oraciones atrás. Esto no excluye que hubiera trabajos e ideas lógicas interesantes sobre las modalidades, y en efecto las podemos encontrar, principalmente en Aristóteles, los megáricos, los estoicos y los filósofos medievales.

Aristóteles escribió mucho sobre modalidades. No dedicó al tema un libro o un trabajo largo, pero lo trató en muchos pasajes de su obra. Por ejemplo, en *Sobre la interpretación,* reflexionó sobre las relaciones entre las modalidades, en *Analíticos Primeros,* construyó una teoría sobre el «silogismo modal» y en *Tópicos* usó las nociones modales en su teoría de la predicación, donde distingue, por ejemplo, entre rasgos que un hombre tiene necesariamente y otros que puede o no tener; en el caso de que efectivamente tenga rasgos del último tipo, son sólo «propiedades accidentales» del hombre en cuestión. Veamos más de cerca ideas suyas conectadas con preocupaciones actuales.

En *Sobre la interpretación* y *Analíticos Primeros,* Aristóteles no intenta dar un análisis ni de la *necesidad* ni de la *posibilidad* pero observa que cada una de ellas es definible en términos de la otra y la negación [véase (*D*1), de la sección anterior]. También señala que lo *contingente* es lo posible que no es necesario. Es consciente, pues, de la interdefinibilidad de las nociones modales que ya hemos hecho notar. En *Analíticos Primeros* hay una teoría sobre el silogismo modal. Suelen llamarse ahora 'asertóricas' las oraciones que no contienen expresiones con un contenido

modal (*expresiones modales,* a partir de ahora). Aristóteles analiza las oraciones que se obtienen insertando expresiones modales en oraciones asertóricas. Un silogismo modal se obtiene de un silogismo asertórico insertando expresiones modales en las premisas (en una o en las dos) y/o en la conclusión. Dificulta el análisis de un texto clásico sobre estas cuestiones el hecho de que la inserción de expresiones modales no produce oraciones inambiguas; por ejemplo, a veces es difícil determinar si se intentó usar modalidades *de dicto* o *de re.* Como la distinción es relevante para evaluar algunas tesis aristotélicas, daré un ejemplo en que se construyen dos oraciones modales, una *de dicto* y otra *de re,* partiendo de una misma oración asertórica, que será de un tipo que aparece con frecuencia en la silogística clásica: una oración universal afirmativa. No daré un significado preciso a las expresiones modales que use, porque el objetivo del ejemplo es mostrar de modo intuitivo la diferencia *de dicto/de re,* y esto puede lograrse aun usando las expresiones-clave de un modo bastante vago (además, usar significados más precisos podría prejuiciar la lectura de Aristóteles).

La oración 'Todos los estudiantes que llegaron tarde a la clase llegaron temprano a la misma' es absurda, si hacemos el supuesto de que las palabras se usan de las maneras habituales y 'la misma' se refiere a la clase mencionada primero. Prefijando a la oración la expresión 'Posiblemente', con la intención de hacer una afirmación modal *de dicto,* obtenemos algo como: 'Posiblemente, todos los estudiantes que llegaron tarde a la clase llegaron temprano a la misma'. Para que la frase total sea verdadera, la oración que sigue al 'Posiblemente', debe ser *posible,* debe existir la posibilidad de que sea verdadera. Pero no existe tal posibilidad, bajo el supuesto hecho acerca del uso de las palabras en la oración (para simplificar, supongo que las proposiciones universales afirmativas se usan de un modo que escapa a la crítica moderna al cuadrado clásico de oposición). La frase total que construimos es, pues, falsa. Tomemos ahora la oración (universal afirmativa) 'Todos los estudiantes que llegaron tarde a la clase *hubieran podido* llegar temprano a la misma'. La expresión subrayada se comporta como una expresión modal usada *de re:* no se afirma ahora que cierta oración *p* es posible; se afirma algo de ciertas personas, no de ciertas palabras. Se dice que ciertos estudiantes hubieran tenido la posibilidad de comportarse de otra manera distinta de como se comportaron: llegaron tarde pero podrían haber llegado temprano. Esta afirmación es más plausible que la afirmación modal *de dicto* que analizamos antes. Prosigamos ahora con las ideas aristotélicas sobre las modalidades.

Desgraciadamente, la teoría aristotélica del silogismo modal es muy confusa, y los autores que la han estudiado en detalle suelen afirmar que contiene errores importantes. Al analizar el silogismo modal que responde a la forma del *Barbara* de primera figura, solo que con las premisas y la conclusión afectadas por expresiones modales conectadas con la idea de posibilidad, Aristóteles lo considera *válido* (si se me permite seguir

usando terminología posterior). Kneale y Kneale (1962, 83) interpretan que las expresiones modales se usan *de dicto* en *Sobre la interpretación* y *Analíticos Primeros*[4]. Si esto es así, es fácil encontrar contraejemplos a la validez del silogismo modal mencionado unas líneas atrás: uno de ellos se construye afectando con palabras que expresan posibilidad *de dicto* las premisas 'Todo triángulo es azul' y 'Toda cosa roja es un triángulo', así como la conclusión que se sigue de ellas según la forma del *Barbara* ya mencionado (adapto un ejemplo de Kneale y Kneale, 1962, 88, donde pueden encontrarse otras ilustraciones; la afirmación aristotélica sobre la validez del caso de silogismo modal analizado puede encontrarse en Aristóteles, 1984, 53 o Aristóteles, 1988, v.2, 140-141)[5]. Llama la atención el contraste entre el estudio del silogismo modal y la teoría del silogismo común, primera teoría lógica que vio el mundo y que Aristóteles desarrolló con gran virtuosismo (los errores son en este caso de poca monta comparados con los hallazgos). Se han dado diversas explicaciones de este hecho extraño, entre ellas la hipótesis de que la teoría del silogismo modal es un agregado tardío y apresurado que hizo Aristóteles a una versión ya acabada de los *Analíticos Primeros*. Para terminar con los ejemplos de ideas aristotélicas acerca de las modalidades, observemos que su teoría de la predicación (que mencionamos antes) es claramente un ejemplo de lo que hoy se entiende por 'esencialismo' (véase sección I). Las modalidades usadas en esta teoría son, pues, *de re*.

Varios autores que han tratado de formular una teoría aristotélica coherente sobre las modalidades (reuniendo sus textos acerca del tema, extrayendo sus consecuencias, etc.), han llegado a la conclusión de que es una misión imposible (van Rijen, 1989, es un libro más optimista al respecto).

Entre los megáricos contemporáneos de Aristóteles se negaba la diferencia entre acto y potencia, y aparentemente esa posición conducía a un rechazo de las distinciones modales. Una generación después, sin embargo, la discusión de las modalidades atrajo a los megáricos y Diodoro Cronos propuso una interesante definición de ellas. No se dispone de sus textos, pero de acuerdo con el testimonio de Boecio, Diodoro definía lo posible como lo que es o va a ser, lo imposible como lo que, siendo falso, no será verdadero, lo necesario como lo que, siendo verdadero, no será falso y lo no-necesario como lo que o bien ya es falso, o lo será (posiblemente Boecio no recogió una referencia a la verdad que figuraba en la definición de lo posible; un texto de Cicerón y el hecho de que los valores de verdad son mencionados en las otras definiciones sugieren esto).

4. Los autores mencionados hacen análisis de expresiones griegas (aunque en general usan la traducción clásica de Ross); lidiando sólo con traducciones, yo no he podido formarme una opinión firme acerca de la interpretación que menciono en el texto.

5. Es un ejercicio instructivo analizar el ejemplo propuesto aplicando primero a las premisas y la conclusión palabras modales que expresan posibilidad *de dicto* y hacer luego lo mismo con palabras que expresan posibilidad *de re*.

Entre otras cosas, llama la atención que Diodoro mezcle las nociones modales con distinciones temporales, como se hace hoy en día en ciertas lógicas; también es curioso que, como en la lógica temporal actual, se conciba que los valores de verdad[6] pueden cambiar con el tiempo. Diodoro sostenía su definición de lo posible con un argumento de cierta complejidad (véase Kneale y Kneale, 1962, 119). Filón de Megara se opuso a Diodoro y, al parecer, defendió una noción de *posibilidad* equivalente a la auto-consistencia (más similar, entonces, a la *posibilidad* carnapiana). Se conocen menos las ideas de los estoicos, pero al parecer eran más similares a las de Filón que a las de Diodoro, aunque hacían algunas críticas al primero (véase Kneale y Kneale, 1962, 123).

El pensamiento medieval fue muy rico en discusiones filosóficas sobre las modalidades. En el siglo XII, el influyente lógico Abelardo distinguió entre proposiciones que atribuyen *modi* (*i.e.,* modos: necesidad, posibilidad o imposibilidad) a *dicta* (*i.e.,* proposiciones) y proposiciones en las que se atribuyen modalmente ciertas características a sujetos que no son *dicta.* Esto sugiere inmediatamente la distinción entre modalidades *de dicto* y *de re;* pero Abelardo cree que las proposiciones modales genuinas son las últimas mencionadas, en las cuales una palabra modal califica el vínculo entre un sujeto y cierta característica que se le atribuye. 'Es posible que Sócrates esté corriendo' no expresa una genuina proposición modal, como sí lo hace 'Sócrates posiblemente está corriendo'. Las opiniones de Abelardo sobre las modalidades influyeron mucho sobre Guillermo de Shyreswood y fueron tomadas en cuenta por Tomás de Aquino. Pero a diferencia de aquellos autores, Tomás de Aquino otorga un *status* modal genuino a las dos proposiciones sobre Sócrates que nos sirvieron antes de ejemplos. Para este filósofo cristiano, las proposiciones modales genuinas pueden ser *de dicto* o *de re.* En su *Summa contra gentiles,* Tomás de Aquino, como muchos otros filósofos durante el siglo siguiente, vinculó esta distinción con el problema de la batalla naval de Aristóteles y algunas objeciones al conocimiento divino de «futuros contingentes». Dada la influencia de Tomás de Aquino en la filosofía posterior, la distinción entre las modalidades *de dicto* y *de re* fue adoptada después por muchos autores.

Los filósofos modernos se han ocupado de las modalidades, pero más en conexión con cuestiones teológicas y metafísicas que en relación con la lógica. Knuuttila (comp., 1988) es una antología útil para tener alguna información de las ideas sobre las modalidades en la filosofía moderna (incluye un artículo sobre el positivismo lógico).

Los temas de la «prehistoria modal» no serán retomados en el resto del trabajo, de modo que será oportuno dar aquí alguna orientación bibliográfica sobre el período. El clásico Kneale y Kneale (1962) es una obra de consulta ideal para comenzar: tiene mucha información histó-

6. ¿De qué entidades? Los documentos no permiten decir con certeza a qué entidades atribuía Diodoro valores de verdad.

rica sobre la lógica modal tomada de las fuentes más directas asequibles y también comentarios interesantes desde un punto de vista lógico moderno (para ubicar las partes relevantes, véanse en el índice temático de la obra las entradas que comienzan con *modal* o *modality*). Naturalmente, por su fecha de publicación no menciona ediciones de autores clásicos posteriores a 1962 ni literatura secundaria que haya aparecido después de esa fecha; pero ahora daré información complementaria. Aristóteles (1984) y Aristóteles (1988) son ediciones recientes de obras aristotélicas (la primera en inglés, la segunda en español) que contienen los tratados lógicos de ese autor a los que nos hemos referido (y en el segundo caso, abundante información sobre ediciones anteriores). Hay una traducción española reciente (y fiable, según me dicen) de la obra de Tomás de Aquino que mencioné en el texto: Tomás de Aquino (1991). El primer intento de formalizar la lógica modal aristotélica se encuentra en el libro de McCall (1963). Dos libros recientes interesantes son van Rijen (1989), sobre la lógica modal de Aristóteles, y Knuuttila (1993), acerca de las modalidades en la filosofía medieval, pero con un capítulo inicial sobre Aristóteles. La profusa bibliografía de esta última obra es muy adecuada para actualizar la información sobre fuentes secundarias posteriores a Kneale y Kneale (1962).

2. *La etapa sintáctica*

La figura más importante de este período fue C. I. Lewis. Su trabajo en lógica modal marcó el comienzo, no sólo de la etapa sintáctica, sino de la historia de esta disciplina en su forma moderna. La publicación del primer volumen de *Principia Mathematica,* de Whitehead y Russell, en 1910, influyó mucho sobre su obra. En la lógica proposicional de los *Principia* se usa el condicional material, también llamado por los autores 'implicación material'. Son derivables en esa lógica, entonces, las llamadas 'paradojas de la implicación material', de las cuales las más conocidas son:

(1) $q \supset (p \supset q)$

(2) $\sim p \supset (p \supset q)$

Puede decirse que (1) expresa la idea de que una proposición verdadera es implicada (materialmente) por cualquier proposición y que (2) afirma que una proposición falsa implica (materialmente) cualquier proposición. Otra paradoja interesante es:

(3) $(p \supset q) \vee (q \supset p)$,

según la cual, dadas dos proposiciones cualesquiera, siempre están «conectadas» por la implicación material: o bien la primera implica (materialmente) la segunda, o la segunda tiene esa relación con la primera.

Lewis no discrepaba con (1)-(3); pensaba que eran auténticas leyes lógicas, dado el significado con que se usaba $\supset$ en *Principia.* Pero el punto de partida de sus investigaciones en lógica modal (en Lewis, 1912), fue la observación de que hay una implicación distinta de la material, más

fuerte que ella, y con diferentes leyes formales (por ejemplo, no tiene las propiedades formales que (1)-(3) asignan a $\supset$). Lewis la llamó 'implicación estricta' y le asignó un símbolo que nosotros reemplazaremos con '$\rightarrow$'. Puede leerse '$p \rightarrow q$' de varias maneras alternativas: p implica estricta, o necesariamente, q, o también: *q se sigue de p.* La última lectura sugiere que la implicación estricta de Lewis es la implicación lógica, y en efecto, $p \rightarrow q$ puede definirse como $\sim \Diamond(p \wedge \sim q)$, si $\Diamond$ se entiende con el sentido de la posibilidad lógica (de modo que para definir '$\rightarrow$' —en el sentido de Lewis— con los operadores modales monádicos, éstos últimos deben usarse como Carnap más bien que como Kripke; esto se debe a que la *implicación lógica* se define en términos de *posibilidad lógica,* no *metafísica: p* implica lógicamente q si y sólo si no es lógicamente posible que p sea verdadero y q falso, de acuerdo con las definiciones usuales). Como '$\rightarrow$' puede definirse usando un operador modal, lo consideraremos también un operador modal (binario); como se aplica a enunciados enteros y no a sus partes, expresa una modalidad *de dicto:* la *implicación lógica* es una relación entre proposiciones, desde el punto de vista usual entre los lógicos.

La *implicación estricta* permite introducir la *equivalencia estricta.* Simbolizando ésta última mediante '$\leftrightarrow$' (apartándonos otra vez de la notación de Lewis), podemos definirla así:

(*D*4) $(p \leftrightarrow q) =_{\text{def.}} (p \rightarrow q) \wedge (q \rightarrow p)$

Es obvio que si en las fórmulas (1)-(3) se reemplaza $\supset$ por $\rightarrow$, no se obtienen leyes lógicas aceptables (por ejemplo, no es cierto que dadas dos proposiciones p y q cualesquiera, se cumpla que p implique estrictamente a q o viceversa). Pero entonces se requeriría un sistema formal distinto del usual para expresar las leyes lógicas de la implicación estricta. Lewis emprendió la tarea y después de una serie de artículos (el primero de los cuales fue Lewis, 1912), publicó un sistema axiomático para la implicación estricta en su libro *A Survey of Symbolic Logic* (Lewis, 1918; la teoría axiomática mencionada se conoce en la literatura como 'el sistema del *Survey*'). Otros autores habían observado ya que había una implicación más fuerte que la material —del tipo de la implicación estricta de Lewis— e incluso habían expresado algunas propiedades de ella mediante un lenguaje simbólico; pero no habían usado el método axiomático para estudiarla y es por eso que no se les reconoce la paternidad de la lógica modal moderna (pienso en Hugh MacColl; no he podido consultar MacColl, 1880, pero por lo que dicen lógicos que conocen este trabajo, su autor estuvo muy cerca de formular por primera vez un sistema de lógica modal). Lewis quiso llenar una laguna de *Principia;* pero lo hizo empleando el método axiomático de esa obra y por eso logró el lugar que tiene en la disciplina que nos ocupa.

Dos sistemas axiomáticos que comparten el mismo lenguaje formal pueden diferir en los conjuntos de axiomas y reglas de inferencia de los cuales parten y tener sin embargo exactamente el mismo conjunto de teoremas. Facilitará mucho mi exposición llamar a veces con el mismo nom-

bre a sistemas así relacionados. Similarmente, dos sistemas pueden diferir porque adoptan distintos símbolos primitivos (en cuyo caso puede considerarse que no comparten el mismo lenguaje formal, si se adopta la idea de que las definiciones introducen abreviaturas en el metalenguaje) y sin embargo puede ocurrir que haciendo caso omiso de cuál es la notación primitiva y cuál la definida no haya diferencia en el conjunto de teoremas. También en casos como éste disimularé a veces las diferencias y emplearé un mismo nombre para sistemas así relacionados[7]. En lo que sigue daré alguna información histórica acerca de los sistemas axiomáticos más conocidos en lógica modal proposicional (en la sección III estudiaremos algunos de ellos en forma sistemática).

En Lewis (1918) se analiza un solo sistema de lógica modal, el ya mencionado sistema del *Survey*. Pero Lewis advirtió en seguida que había cuestiones sobre las cuáles el sistema del *Survey* no tomaba partido: por ejemplo, en el sistema no se establece si □ p y □□p son equivalentes o no (algunos autores parecen creer que el sistema mencionado se pronuncia negativamente sobre la equivalencia aludida, pero eso no es exacto: si así fuera, S4 no podría ser una extensión consistente del sistema del *Survey;* en este último sistema no es teorema que las fórmulas mencionadas sean equivalentes pero *tampoco* es teorema la negación de esa afirmación). Lewis advirtió también que nuestras intuiciones sobre las nociones modales no se pronuncian claramente sobre cuestiones como la mencionada, y otras similares. En otras palabras, comprendió que había más de un sistema modal plausible. En Lewis y Langford (1932) se analizan cinco sistemas modales y se les da el nombre por el cual se los conoce desde entonces: S1, S2, S3, S4 y S5. S1 y S2 se desarrollan en detalle en el libro; en un apéndice se analizan más brevemente los tres restantes. S3 coincide con el sistema del *Survey* (haciendo caso omiso de diferencias como las mencionadas en el párrafo anterior). Daré alguna información sobre características importantes de los cinco sistemas.

Los cinco sistemas carecen de una semántica sistemática; sus primitivos se aclaran informalmente, más o menos de la manera en que lo hemos hecho aquí. Los cinco sistemas son extensiones de la lógica proposicional veritativo-funcional clásica (LPC, en lo que sigue); pero en la formulación de Lewis ninguno es una extensión de LPC de manera explícita, *i. e.*, ninguno es el resultado de agregar axiomas a un sistema formal que expresa de manera completa LPC. Después del primero, cada sistema es una extensión propia del sistema anterior (*propia* en el sentido de que hay un auténtico agregado de nuevos teoremas). Si llamamos 'leyes reductivas' a teoremas que muestran la equivalencia de concatenaciones de operadores modales de distinta longitud (por ejemplo, □ y □□; en la sección III veremos con más detalle qué es una ley reductiva), S1 y S2 carecen

7. En los textos de lógica modal es bastante habitual que los autores se comporten como si siguieran esta convención (por ejemplo, llamando 'S4' a más de un sistema, todos ellos difiriendo en las formas descriptas en el texto), pero sin hacerla explícita.

de leyes reductivas, S3 las tiene (como se probó en Parry, 1939) y también S4 y S5. De hecho, S4 y S5 se obtienen agregando axiomas sugeridos por Becker (1930) y que tienen leyes reductivas como consecuencias inmediatas.

S4 y S5 son dos de los tres sistemas modales más estudiados. El tercero es T, cuya historia resumimos a continuación. Gödel (1933) fue el primer trabajo en que un sistema modal se obtuvo mediante una extensión explícita de LPC. Quitando un axioma de la formulación gödeliana de ese sistema, Feys (1937) obtuvo el sistema que ahora llamamos 'T'. S1 y S2 son más débiles que T, que a su vez es más débil que S4 y S5. T y S3 son incomparables (ninguno de los dos sistemas es una extensión del otro).

La segunda figura más importante en la etapa sintáctica de la lógica modal es von Wright. Von Wright (1951) introdujo los sistemas modales M, M' y M". Sobociński (1953) mostró que estos sistemas son respectivamente equivalentes a T, S4 y S5 (sistemas que analizaremos en la sección III).

3. *La etapa semántica*

La presente sub-sección y la siguiente serán muy breves porque los temas se prestan más a un tratamiento sistemático que histórico, y serán objeto de las secciones IV y V. Sólo indicaré algunos datos que permitan ubicar los desarrollos y los protagonistas en el tiempo.

Hemos fijado el comienzo de la etapa semántica en 1959. Sin embargo, ya en Carnap (1947) se había construido una semántica para la lógica modal, inspirada en la idea de Leibniz según la cual algo es necesariamente verdadero cuando es verdadero en todo mundo posible. Carnap introducía ciertos conjuntos de fórmulas que llamaba 'descripciones de estado' y que cumplían el papel de los mundos posibles en su semántica (véase Carnap, 1947, 9). También definía la noción de 'valer (*hold*) en una descripción de estado dada', que era en la teoría el análogo de 'ser verdadero en un mundo posible' (Carnap, 1947, 9). Con estos elementos, Carnap pudo definir las condiciones de verdad de fórmulas de la estructura $\Box p$, traduciendo a su lenguaje de descripciones de estado la idea de Leibniz (para eso deben combinarse las definiciones 2-2 de Carnap, 1947, 10 y 39-1 de Carnap, 1947, 174). En términos modernos, esto significa que Carnap disponía de una definición de *verdad* para el lenguaje de LPC enriquecido con $\Box$. Contando con tal definición, se puede definir también la noción de *fórmula válida* para tal lenguaje, cuantificando sobre descripciones de estado (véase Bull y Segerberg, 1984, 13, para ver con más detalle cómo se puede introducir de esta manera la noción de *validez*). Las fórmulas que resultan válidas de acuerdo con la definición son exactamente los teoremas de S5. En otras palabras, y empleando nuevamente terminología actual: Carnap construyó una semántica adecuada para S5. ¿Por qué no pensar entonces que la etapa

semántica comenzó en 1947? La respuesta está vinculada con rasgos de la semántica modal posterior. A partir de Kripke (y otros autores que mencionaremos), se pudo disponer de métodos con los cuales pueden construirse semánticas adecuadas para una gran diversidad de sistemas modales (por ejemplo, para T, S4 y S5). La semántica de Carnap no tenía esa flexibilidad. Desde la óptica contemporánea, se piensa que proporcionó una semántica para un sistema modal aislado, y no para la lógica modal.

Alrededor de doce años más tarde, tres lógicos descubrieron (al parecer de manera independiente) los métodos semánticos mencionados hacia el final del párrafo anterior. Los históricos trabajos en que expusieron sus resultados son Kanger (1957), Kripke (1959), Kripke (1963a), Kripke (1963b), Hintikka (1961) y Hintikka (1963). Con estos trabajos nació la moderna semántica de los mundos posibles. Los métodos de los tres autores tienen una semejanza muy profunda, aunque los detalles de presentación difieran. Los trabajos de Kripke fueron los que alcanzaron mayor difusión y por eso adoptamos el año 1959 (cuando publicó su primer artículo sobre estos temas) como señal del comienzo de la etapa semántica. Un rasgo de las investigaciones que tuvieron lugar en esta etapa fue el interés por estudiar propiedades metalógicas (particularmente la completitud) de sistemas que se habían desarrollado antes de manera sintáctica. En la sección IV veremos cómo se aplican los métodos semánticos de Kripke a los sistemas T, S4 y S5.

4. *La época de la metalógica modal generalizada*

Si uno compara superficialmente un manual como Hughes y Cresswell (1968) con Hughes y Cresswell (1984), o Jansana (1990), la primera diferencia que salta a la vista es que el interés teórico se ha desplazado del análisis de sistemas modales particulares al estudio de grandes familias de tales sistemas. Es difícil asignar una fecha al comienzo de este período, porque el interés por problemas y resultados más generales fue creciendo gradualmente (hasta que el perfil de las investigaciones cambió por completo). Como en el caso de los dos períodos anteriores, adoptaremos un trabajo importante como señal del comienzo de una etapa: Lemmon y Scott (1977). Pero esta publicación tuvo lugar una década después de la muerte de Lemmon, y en realidad la obra había tenido influencia desde ocho o nueve años antes de ser impresa (los lógicos modales usaban copias de las notas de Lemmon en que se basó el libro). Consideraremos pues, que hacia fines de la década de los años 60 comenzó nuestro período. Lemmon y Scott (1977) muestra un claro interés por alcanzar resultados generales sobre sistemas modales, particularmente acerca de completitud. Otros trabajos muy importantes de este período son Segerberg (1971), Goldblatt (1976) y Van Benthem (1982). Daré más información sobre estas investigaciones en la sección V.

III. T, S4 Y S5: ESTUDIO SINTÁCTICO

T, S4 y S5 son los sistemas modales proposicionales más conocidos. Los analizaremos después de introducir ciertas nociones generales.

1. *El lenguaje modal L. Modalidades normales y leyes reductivas*

En nuestra formulación, T, S4 y S5 compartirán un mismo lenguaje, que llamaré 'L'. El vocabulario primitivo de L está formado por los paréntesis (,), las variables proposicionales *p, q, r,* eventualmente usadas con subíndices, los operadores monádicos $\sim$, $\Box$ y el operador binario $\supset$; los operadores $\Diamond$, $\wedge$, $\vee$, $\equiv$, $\rightarrow$, $\leftrightarrow$ (monádico el primero, binarios los restantes) se definirán en términos del vocabulario primitivo. Las reglas de formación son las usuales en lógica proposicional veritativo-funcional, sólo que ahora enriquecidas por la presencia de los operadores modales. Letras mayúsculas como *A, B,* etc., serán usadas como variables metalógicas de fórmulas. Los conectivos $\wedge$, $\vee$, $\equiv$ se introducen con cualquiera de las definiciones usuales. En nuestra formulación de T, S4 y S5, $\Diamond$, $\rightarrow$ y $\leftrightarrow$ se definen así:

(*DL*1) $\Diamond A =_{\text{def.}} \sim\Box\sim A$

(*DL*2) $(A\rightarrow B) =_{\text{def.}} \Box(A\supset B)$

(*DL*3) $(A\leftrightarrow B) =_{\text{def.}} ((A\rightarrow B)\wedge(B\rightarrow A))$

Cualquier sistema axiomático cuyo lenguaje sea L será considerado un *sistema modal.* Llamaré *modalidad* a cualquier secuencia finita O_1, O_2, ..., O_n, de operadores monádicos de L, donde cada O_i es, entonces, o $\sim$, o $\Box$ o $\Diamond$; *n* es la longitud de la modalidad. Por cuestiones de simetría, también se llamará *modalidad* a la secuencia nula, que es una serie de operadores monádicos que tiene longitud 0. Llamaremos *modalidad normal* a una modalidad que no contiene ningún $\sim$ o bien contiene uno solo y al comienzo de la secuencia. Una modalidad normal se dice *iterada* si no es nula y más de un lugar de la secuencia es ocupado por un operador distinto de $\sim$. Por ejemplo, $\Box\sim$, $\sim\Box$ y $\sim\Box\Diamond$ son modalidades, pero la primera no es normal, como sí lo son la segunda y la tercera; sólo la última es iterada. Una *ley reductiva* de un sistema modal S es un teorema de S que tiene la forma (M$p\equiv$Np), donde M y N son modalidades normales de distinta longitud; si S posee tal teorema, se dirá que M y N son modalidades *equivalentes* en S (y diremos también que la de mayor longitud *se reduce* a la otra). Si M es una modalidad normal y en S no hay ninguna modalidad equivalente a M y de longitud menor, diremos que M es una modalidad *irreducible* de S.

2. *El sistema T*

Seguiré de cerca la presentación de Hughes y Cresswell (1968), aunque con algunos cambios. Uno de ellos es el reemplazo del subsistema PM, con el que ellos expresan la lógica proposicional veritativo-funcional, por

el sistema axiomático con que Mendelson (1987) formaliza esa parte de la lógica (aunque usaré una regla de substitución en lugar de los axiomas-esquema de Mendelson).

Ya vimos el lenguaje y las definiciones que se usan en T. Los axiomas son los siguientes:

(A1) $(q \supset (p \supset q))$

(A2) $[(p \supset (q \supset r)) \supset ((p \supset q) \supset (p \supset r))]$

(A3) $[(\sim q \supset \sim p) \supset ((\sim q \supset p) \supset q)]$

(A4) $\Box p \supset p$

(A5) $\Box(p \supset q) \supset (\Box p \supset \Box q)$

Los tres primeros axiomas son tautologías de la lógica proposicional veritativo-funcional clásica (LPC). El cuarto axioma expresa la idea de que si una proposición es necesaria, es verdadera. Obsérvese que de acuerdo con (*DL*2), la implicación estricta se puede definir con el antecedente de (A5). En ese caso, se podría traducir la idea formalizada por (A5) así: si una proposición p implica una proposición q, se cumple que si p es necesaria, q también lo es. (A5) transmitiría entonces la vieja idea de que las proposiciones necesarias sólo implican lógicamente proposiciones necesarias.

T tiene tres reglas de inferencia:

(RS) *Regla de substitución uniforme:* Si en un teorema[8] se substituye una variable proposicional por una fórmula bien formada, de manera uniforme (*i.e.*, todas las ocurrencias de la variable se reemplazan siempre por la misma fórmula), el resultado también es un teorema.

(MP) *Modus ponens:* De un condicional material y su antecedente, se puede derivar su consecuente.

(RN) *Regla de necesidad:* Si una fórmula es teorema, el resultado de prefijarle el operador de necesidad también es un teorema. Si $\vdash A$ significa que A es teorema (del sistema del cual se esté hablando; en caso de ambigüedad, se aclarará cuál es el sistema al que se desea hacer referencia), y $\vdash A \Rightarrow \vdash B$ se usa para indicar que si A es teorema, puede derivarse sintácticamente que B también lo es, (RN) puede formularse así:

$\vdash A \Rightarrow \vdash \Box A$

Los tres primeros axiomas, junto con las dos primeras reglas, forman el sistema que Mendelson llama 'L_3', si se quitan del lenguaje los operadores modales. Puede mostrarse fácilmente que L_3 tiene los mismos teoremas que el sistema que Mendelson llama 'L' (Mendelson, 1987, 39, ejercicio 1.55), para el cual se ha demostrado que su conjunto de teoremas coincide con el de las tautologías (Mendelson, 1987, 33-35). En ese caso, T contiene como teoremas todas las tautologías y los ejemplos de ellas en que las variables proposicionales son reemplazadas (uniformemente) por fórmulas de nuestro lenguaje L.

Es fácil establecer que la siguiente es una regla derivada:

(R4) $\vdash A \supset B \Rightarrow \vdash \Box A \supset \Box B$

8. Usamos la palabra 'teorema' en el sentido en que abarca también a los axiomas.

Formularemos algunos teoremas importantes para ver qué contiene T. Para que se advierta el estilo en que pueden demostrarse teoremas en nuestra formulación del sistema, haremos algunas demostraciones explícitas; luego las abreviaremos o suprimiremos, pero un lector con alguna práctica en LPC podrá completarlas, si comprendió los primeros ejemplos. Una enunciación del teorema precederá a su demostración (cuando se acompañe una demostración). La notación p/A indicará que la variable p se debe substituir por la fórmula A (de manera uniforme)[9].

T1: $p \supset$

(1) $\Box \sim p \supset \sim p$	(RS en A4, $p/\sim p$)
(2) $\sim \sim p \supset \sim \Box \sim p$	((1), Transposición)
(3) $p \supset \Diamond p$	(Doble Negación, (2), Silogismo Hipotético y $DL1$)

T2: $\Box(p \wedge q) \equiv (\Box p \wedge \Box q)$

(1) $(p \wedge q) \supset p$	(Tautología)
(2) $(p \wedge q) \supset q$	(Tautología)
(3) $\Box(p \wedge q) \supset \Box p$	((1), R4)
(4) $\Box(p \wedge q) \supset \Box q$	((2), R4)
(5) $\Box(p \wedge q) \supset (\Box p \wedge \Box q)$	

(5) se sigue de (3) y (4) aplicando la tautología $(p \supset q) \supset [(p \supset (p \supset (q \wedge r))]$.

Obsérvese que (5) es la «mitad» de T2. Probemos la otra mitad.

(6) $p \supset (q \supset (p \wedge q))$	(Tautología)
(7) $\Box p \supset \Box(q \supset (p \wedge q))$	((6), R4)
(8) $\Box(q \supset (p \wedge q)) \supset (\Box q \supset \Box(p \wedge q))$	(RS en A5)
(9) $\Box p \supset (\Box q \supset \Box(p \wedge q))$	(7 y 8, por Silogismo Hipotético)
(10) $(\Box p \wedge \Box q) \supset \Box(p \wedge q)$	((9), Exportación y MP)

Obsérvese que ésta es la mitad que nos faltaba del T2. Poniendo en conjunción (5) y (10) y aplicando la definición usual de $\equiv$ se obtiene:

(11) $\Box(p \wedge q) \equiv (\Box p \wedge \Box q)$

Para LPC vale el metateorema de la equivalencia: si B se obtiene de A reemplazando una o más ocurrencias de C por D y resulta que $\vdash C \equiv D$, entonces B es un teorema, si A lo era.

Es fácil demostrar que

$\vdash (A \equiv B) \Rightarrow \vdash (\Box A \equiv \Box B)$

es una regla derivada de T. También es fácil probar entonces que el metateorema de la equivalencia puede extenderse a T. Usaremos este resul-

9. Presupongo el conocimiento de varias tautologías y reglas de LPC, así como de sus nombres habituales. Si se desea explicitar completamente una prueba, debe evitarse el reemplazo de equivalentes (salvo que lo sean por definición) antes del T_3 (por ejemplo, en el paso 3 de la prueba de T_1 debe usarse $(p \supset \sim \sim p)$ y no la equivalencia correspondiente y un reemplazo permitido por ella).

tado en las próximas demostraciones, indicando el uso del metateorema mediante 'MRE' (metateorema del reemplazo de equivalentes).

T3: $\Box p \equiv \sim \Diamond \sim p$

(1) $\sim \Diamond \sim p \equiv \sim \Diamond \sim p$	(ejemplo de tautología)
(2) $\sim \sim \Box \sim \sim p \equiv \sim \Diamond \sim p$	(*DL*1)
(3) $\Box p \equiv \sim \Diamond \sim p$	(Doble Negación, MRE)

Son obvios corolarios de T3:

T4: $\sim \Box p \equiv \Diamond \sim p$

T5: $\sim \Diamond p \equiv \Box \sim p$

Usando MRE y los últimos corolarios puede demostrarse que en T toda modalidad puede «normalizarse»: dada una modalidad M cualquiera, existe una modalidad normal M', de igual o menor longitud que M, y tal que es teorema de T la fórmula ($Mp \equiv M'p$). Este resultado es obvio para la modalidad nula y las de longitud 1. Por inducción matemática se generaliza fácilmente para modalidades de longitud $n+1$, para cualquier n positivo (dada una modalidad de longitud $n+1$, separe su primer componente y aplique la hipótesis inductiva a la cadena de los n restantes; luego divida en cuatro casos las combinaciones posibles que se pueden dar entre el componente que había quedado aislado y el primer componente de la modalidad ya normalizada: negación/negación, operador modal/negación, etc.).

Otro teorema importante de T (fácilmente demostrable usando MRE y los resultados ya establecidos), es:

T6: $\Diamond(p \vee q) \equiv (\Diamond p \vee \Diamond q)$

El teorema 2 y el 6 muestran, respectivamente, que la necesidad es distributiva respecto de la conjunción y la posibilidad respecto de la disyunción. En cambio, la necesidad no es distributiva respecto de la disyunción ni la posibilidad respecto de la conjunción. En los dos últimos casos, sólo es teorema una «mitad» de lo que sería la ley distributiva. Son los teoremas:

T7: $(\Box p \vee \Box q) \supset \Box(p \vee q)$

T8: $\Diamond(p \wedge q) \supset (\Diamond p \wedge \Diamond q)$

Un rasgo interesante de T es que en el sistema se pueden demostrar unas «paradojas de la implicación estricta». No se puede demostrar que una verdad es implicada por cualquier proposición, ni que una falsedad implique lógicamente cualquier proposición, pero si cambiamos *verdad* por *verdad necesaria* y *falsedad* por *imposibilidad* obtenemos leyes válidas de la implicación estricta. Formalmente, tenemos:

T9: $\Box q \supset (p \rightarrow q)$

T10: $\sim \Diamond p \supset (p \rightarrow q)$

Hay algunos rasgos metalógicos de T que es importante destacar.

T es consistente. T es una extensión propia y conservativa de LPC (*i.e.*, agrega teoremas a los de LPC, pero no hay teoremas nuevos formulados únicamente con símbolos de LPC).

T no contiene leyes reductivas. Dada una modalidad iterada cualquiera, no es posible encontrar otra más corta y equivalente a la primera en el sistema T (recuérdese que las modalidades iteradas son una subclase de las normales). Hay en T infinitas modalidades normales no equivalentes entre sí. (Estos hechos acerca de T no se pueden demostrar de manera muy sencilla.)

Para terminar con las observaciones acerca de T, podemos hacer notar que se trata de un sistema bastante intuitivo. En general, los puntos de partida de T no violentan nuestras intuiciones y entre los autores que no tienen objeciones de principio en contra de la empresa misma de construir una lógica modal, no suele haber resistencia ante los axiomas, reglas y definiciones de T. Una excepción es la lógica relevante. Muchos lógicos relevantes piensan que si $\rightarrow$ representa la implicación lógica, la definición que dimos de este operador binario es incorrecta: el *definiens* sólo proporciona condiciones necesarias, pero no suficientes, para una caracterización del *definiendum* (véase Anderson y Belnap, 1975).

3. *El sistema S4*

Una manera sencilla de obtener S4 a partir de nuestra formulación de T es agregar el siguiente axioma:

(A6) $\Box p \supset \Box\Box p$

Salta a la vista que (A6) permite derivar infinitas leyes reductivas. En efecto, como el condicional inverso del axioma introducido es un ejemplo de substitución de (A4), se deriva rápidamente:

T1: $\Box p \equiv \Box\Box p$,

que es una ley reductiva de la cual pueden obtenerse infinitas leyes reductivas adicionales. Por ejemplo, (T1) permite obtener por substitución todas las equivalencias de la forma M$p \equiv$ M'p, donde M y M' son secuencias finitas de $\Box$ y M' tiene la longitud de M más 1. Como es obvio, en S4 sigue valiendo (MRE) que, combinado reiteradamente con ejemplos de substitución de (T1), permite reemplazar en una fórmula cualquiera una secuencia de $n+1\Box$ por un solo $\Box$.

También se infiere rápidamente que la reiteración de $\Diamond$ se comporta como la repetición de $\Box$. En efecto, negando ambos lados del $\equiv$ del (T1) y substituyendo p por su negación obtenemos una fórmula que por aplicación de teoremas y definiciones de T arroja otra ley reductiva importante:

T2: $\Diamond p \equiv \Diamond\Diamond p$

Un poco menos previsibles son los teoremas:

T3: $\Box\Diamond p \equiv \Box\Diamond\Box\Diamond p$

T4: $\Diamond\Box p \equiv \Diamond\Box\Diamond\Box p$,

aunque su demostración no es mucho más complicada que la de los precedentes. Obsérvese que T3 y T4 muestran que la repetición de cualquiera de los bloques $\Box\Diamond$ y $\Diamond\Box$ se comporta como la reiteración de $\Box$ o $\Diamond$. También puede observarse que estos cuatro teoremas de S4 hacen innecesario usar secuencias de operadores monádicos diferentes de $\sim$ y de longitud mayor que 3 (el lector puede tomar una secuencia de tres operadores distintos de $\sim$ y observar si es irreductible el resultado de anteponerle $\Box$ o $\Diamond$). Estas observaciones muestran que toda modalidad normal es reducible en S4, o bien a una de las modalidades de la lista

modalidad nula, $\Box$, $\Diamond$, $\Box\Diamond$, $\Diamond\Box$, $\Box\Diamond\Box$, $\Diamond\Box\Diamond$,

o bien a la negación de una de tales modalidades. Las catorce modalidades mencionadas (las de la lista más sus negaciones) son irreducibles en el sistema que estamos analizando (aunque mostrar esto es más trabajoso que probar que toda otra modalidad normal es equivalente a una de ellas en S4).

S4 es consistente y es una extensión propia y conservativa de LPC. Es también una extensión de T, propia pero no conservativa, ya que en S4 se agregan teoremas que no estaban en T pero son formulables en su lenguaje.

El agregado que se hace en S4 a T no es trivial, en el sentido de que no es intuitivamente obvia la verdad del axioma 6. Si abandonamos el plano intuitivo y construimos una semántica sistemática para los operadores modales, el axioma 6 puede resultar válido o inválido, según la semántica elegida.

4. *El sistema S5*

S5 es una extensión de S4, pero no lo formularemos como una extensión explícita de ese sistema, sino como una extensión de T obtenida por el agregado del axioma:

(A7) $\Diamond p \supset \Box\Diamond p$

Este agregado es menos intuitivo aún que el axioma (6) de S4: lo que afirma es que si una proposición es posible, es necesario que sea posible. Analizaremos las consecuencias de esta adición a T. En las pruebas que siguen usaremos teoremas ya demostrados de T, pero no teoremas de S4, porque aún no se ha mostrado que el presente sistema sea una extensión de aquél.

T1: $\Diamond p \equiv \Box\Diamond p$

(Este teorema es obvio porque el condicional de izquierda a derecha es (A7) y el condicional de derecha a izquierda es un ejemplo de substitución del axioma 4 de T).

T2: $\Diamond \Box p \supset \Box p$

(1)$\Diamond \sim p \supset \Box \Diamond \sim p$ (RS en (A7), $p/\sim p$)

(2)$\sim \Box p \supset \sim \Diamond \Box p$

(Se obtiene del paso anterior por MRE, aplicando dos veces T4 y una vez T5, del sistema T)

(3) $\Diamond \Box p \supset \Box p$ (Uso de Transposición en (2))

T3: $\Box p \equiv \Diamond \Box p$

Este teorema se demuestra mediante dos condicionales: el primero es un ejemplo de substitución de T1 del sistema T (reemplácense las dos ocurrencias de p por $\Box$p) y el segundo es el teorema 2 de S5 que demostramos antes.

T4: $\Box p \supset \Box \Box p$

(1) $\Box p \supset \Diamond \Box p$ (ejemplo del T1 de T)

(2) $\Box p \supset \Box \Diamond \Box p$

(Se obtiene de (1) reemplazando el $\Diamond$ del lado derecho por una combinación equivalente según T1).

(3) $\Box p \supset \Box \Box p$

(Se obtiene de (2) reemplazando el $\Diamond \Box$ del lado derecho por $\Box$, como permiten MRE y T3).

De acuerdo con este teorema, el agregado que hicimos a T para obtener S5 implica el axioma 6 de S4. Se sigue, pues, que S5 contiene todos los teoremas de S4 (recuérdese cómo se obtuvo S4 a partir de T). En particular contiene

T5: $\Box p \equiv \Box \Box p$

T6: $\Diamond p \equiv \Diamond \Diamond p$

Los teoremas 1, 3, 5 y 6 son leyes reductivas que en conjunto son mucho más poderosas que las de S4. Obsérvese que, dada una secuencia cualquiera de dos operadores monádicos distintos de $\sim$, las leyes demostradas permiten suprimir el primero de ellos. Este resultado implica que en S5 todas las modalidades normales pueden reducirse a seis: la nula,

□, ◇ y las negaciones de ellas (y las seis son irreductibles, aunque no lo probaremos aquí). Es un poco sorprendente que todas estas reducciones son consecuencias del agregado de un solo axioma, el 7 de S5. Se puede demostrar un resultado aún más fuerte acerca de S5: en este sistema toda fórmula es equivalente a alguna fórmula en la cual no hay operadores modales aplicados a fórmulas que ya tengan operadores modales.

S5 es consistente. Es una extensión propia y conservativa de LPC y también es una extensión (propia pero no conservativa) de T y S4.

(Si se desea profundizar en los aspectos sintácticos de T, S4 y S5, se pueden consultar las fuentes originales mencionadas en la sección II [la parte sobre «la etapa sintáctica»] o el manual de Hughes y Cresswell, 1968).

IV. T, S4 Y S5: ESTUDIO SEMÁNTICO

En esta sección explicaremos cómo se aplican los métodos de Kripke a los sistemas T, S4 y S5. Pero antes presentaremos una semántica más sencilla, cuyo análisis nos permitirá advertir las motivaciones que tuvo Kripke (o los otros dos pioneros mencionados antes) para introducir algunos elementos que no se habían utilizado hasta entonces en la semántica de sistemas lógicos.

1. *Necesidad leibniziana*

Se atribuye a Leibniz una concepción de la *necesidad* que fue un punto de partida muy fructífero para la semántica modal de nuestros días. Leibniz suponía que había una infinidad de mundos posibles (entre los cuales había elegido Dios el mundo real). Hay proposiciones que son verdaderas en algunos mundos posibles y falsas en otros, pero también hay proposiciones que son verdaderas en *todos los mundos posibles:* son las proposiciones *necesarias.* Por ejemplo, hay mundos posibles (entre ellos el real) en que existen gatos, y mundos posibles en que no existen tales animales; pero en todo mundo posible es verdad que *existen gatos o no existen gatos,* y esa proposición es, entonces, necesaria.

La concepción leibniziana de la necesidad se puede elaborar matemáticamente para definir una noción de *validez* aplicable a las fórmulas de la lógica modal. Cuando se construye una semántica para un cálculo lógico es usual definir las fórmulas válidas como aquellas que son verdaderas en todas las estructuras de cierta familia (véase en Quesada, 1991, cuáles son las estructuras relevantes para la lógica proposicional y cuantificacional de primer orden). Se empieza por definir cuáles son las *estructuras* de la semántica para ese cálculo (son, esencialmente, ciertas asignaciones de entidades a símbolos del cálculo). Se define luego la noción de *verdad respecto de una estructura* (una especificación —usualmente recursiva— del conjunto de fórmulas que resultan verdaderas dadas las

asignaciones hechas en la estructura). Finalmente, se define *fórmula válida* como aquella que es *verdadera en toda estructura.* Podemos construir una semántica de este tipo sobre la base de la concepción leibniziana. Una complicación que se presenta es que la noción de *verdad* deja de ser una relación binaria que conecta fórmulas con estructuras. La razón es que según la idea de Leibniz una proposición puede ser verdadera en un mundo posible y falsa en otro. Cuando tratamos de expresar esta idea en una semántica inspirada en Leibniz, debemos contemplar la posibilidad de que, aun dentro de la misma estructura, una fórmula pueda ser verdadera en un mundo posible y falsa en otro. Construiremos una semántica de este tipo a continuación. Un objetivo de nuestras construcciones será elaborar semánticas *adecuadas* para distintos sistemas lógicos. Consideraremos que una semántica es *adecuada* para un sistema cuando las fórmulas que resultan válidas de acuerdo con esa semántica coinciden con los teoremas del sistema en cuestión.

Simplificará la exposición trabajar con un conjunto fijo de mundos posibles. Supondremos que hay una cantidad infinita denumerable de mundos posibles (*i.e.,* tantos mundos como enteros positivos), que llamaremos m_1, m_2, m_3,... etc. Nuestro lenguaje L consta de infinitas variables proposicionales (*p, q, r,* y las mismas letras con cualquier subíndice entero positivo). Serán estructuras de nuestra semántica aquellas asignaciones que den un valor de verdad unívoco (la verdad o la falsedad) a cada par ordenado de una variable proposicional y un mundo posible (*i.e.,* una estructura se comporta como una función binaria *f* que a cada par de una variable proposicional *v* y un mundo posible *m* le da un único valor de verdad).

Definiremos recursivamente la noción de *verdad* en un mundo en una estructura. La afirmación de que la fórmula *A* es verdadera en el mundo *m* y la estructura *E* se abreviará con la expresión '$E\ (A,\ m) = \mathrm{V}$' (lo cual sugiere —como realmente ocurre— que la definición de *verdad* da lugar a una extensión de la función binaria en que consiste una estructura).

(i) Si *v* es una variable proposicional, *m* un mundo posible y *E* una estructura, $E(v,m) = \mathrm{V}$ si y sólo si (syss, en adelante) *E* le asigna la verdad a *v* en *m*.

Si *m* es un mundo posible y *A, B* son fórmulas,

(ii) $E(\sim A,m) = \mathrm{V}$ syss $E(A,m) \neq \mathrm{V}$

(iii) $E(A \supset B,m) = \mathrm{V}$ syss $E(A,m) \neq \mathrm{V}$ o $E(B,\ m) = \mathrm{V}$

(A partir de (i)-(iii) y las definiciones de los otros conectivos veritativo-funcionales se deducen las cláusulas de la definición de verdad para ellos).

(iv) $E(\Box A,m) = \mathrm{V}$ syss para todo mundo *n,* $E(A,n) = \mathrm{V}$

La última cláusula recoge la idea central de la semántica leibniziana. De ella y las definiciones de los otros operadores modales se deducen las cláusulas para ellos. En particular, se infiere que

(v) $E(\Diamond A,m) = \mathrm{V}$ syss para algún mundo *n,* $E(A,n) = \mathrm{V}$

Como es usual, decir que *A* es *falsa* (en *m,* en *E*) es afirmar que no es verdadera (en *m,* en *E*).

Por último, *A es válida* syss es verdadera en todo mundo de toda estructura.

Veamos cómo se comportan distintas fórmulas respecto de esta última definición. Consideremos

(1) $p \supset \Box p$

(2) $\Box p \supset p$

(3) $\Box(p \supset q) \supset (\Box p \supset \Box q)$

(4) $\Box p \supset \Box\Box p$

(5) $\Diamond p \supset \Box\Diamond p$

(1) es intuitivamente no-válida. También lo es de acuerdo con nuestra semántica: basta considerar una estructura *E* que asigne la verdad a *p* en un mundo *m* y la falsedad en un mundo *n; E* le asigna la verdad al antecedente de (1) en *m* y la falsedad al consecuente de (1) en ese mismo mundo (aplíquense las cláusulas (i), (iv) y (iii)).

(2) es intuitivamente válida y también lo es de acuerdo con nuestra semántica, lo cual se puede probar por el absurdo. Suponer que (2) no es válida implica que hay un mundo *m* y una estructura *E* en que (2) es falsa. En tal caso, su antecedente es verdadero y su consecuente falso en tal mundo. Pero si el antecedente es verdadero en *m,* de acuerdo con (iv), *p* debe ser verdadero en todo mundo, incluido *m,* lo cual contradice la conclusión a la que se había llegado acerca del consecuente. De manera muy similar, se puede probar que (3) es válida.

Es más interesante el caso de (4) porque se trata de una fórmula que no es teorema de T pero sí de S4 y S5. ¿Por qué sistema(s) se decidirá nuestra semántica? Por los dos últimos. Supongamos que el antecedente de (4) es verdadero y su consecuente falso en algún mundo *m* y estructura *E.* Si el antecedente es verdadero en *m, p* es verdadero en todo mundo en *E.* Pero en ese caso, $\Box p$ no solamente es verdadero en *m* sino en todo mundo (en *E*). Si ése es el caso, $\Box\Box p$ también será verdadero en todo mundo en *E* y en particular en *m,* contradiciendo lo supuesto acerca del consecuente. También (5) es válida y la demostración de este hecho es completamente similar a la del anterior.

Nuestra semántica se ha inclinado hacia S5: ha declarado válidas cinco fórmulas que son teoremas de este sistema y entre ellas hay fórmulas que no son teoremas de uno, o dos, de los sistemas anteriores. Se puede probar un metateorema que muestra la adecuación de nuestra semántica al sistema S5: una fórmula es teorema de ese sistema syss es válida de acuerdo con nuestra semántica. Pero entonces nuestra semántica no es adecuada para los sistemas más débiles T y S4. ¿Cómo podría diseñarse una semántica que sólo validara los teoremas de T, o sólo los de S4? Kripke y otros autores ya mencionados introdujeron ideas que permitieron llevar a cabo la empresa y que explicaremos a continuación.

Supongamos que la idea de *mundo posible* es relativa a mundos: un mundo puede ser posible para algunos mundos y no para otros. Emplearemos la siguiente terminología técnica: diremos que el mundo *n* es accesible al mundo *m* cuando *n* es posible para *m.* Abreviaremos mediante

«mRn» la afirmación de que n es accesible a m. Supongamos que la verdad de $\Box A$ en m no requiere la verdad de A en todo mundo posible, sino sólo en aquellos mundos posibles accesibles a m. (iv) y (v) quedarían entonces reformuladas de la siguiente manera:

(iv') $E(\Box A,m)=$ V syss para cualquier n tal que mRn, $E(A,n)=$ V

(v') $E\Diamond(A,m)=$ V syss para algún n tal que mRn, $E(A,n)=$ V.

Hemos introducido una relación entre mundos posibles y hemos redefinido las condiciones de verdad de $\Box A$ y $\Diamond A$ usando esa relación. Pero sólo sabemos de R que es una relación entre mundos. Si no añadimos algo más, no obtendremos una semántica útil. Podríamos agregar postulados acerca de esa relación: que R es reflexiva, que es una relación de equivalencia, etc. Lo que descubrió Kripke es que si trabajamos con (iv') y (v') y agregamos postulados respecto de R, podremos obtener semánticas adecuadas para distintos sistemas si elegimos postulados apropiados para cada uno. Mostraremos cómo se hace esto para T, S4 y S5.

2. *Una semántica para T*

Para dar forma matemática a las ideas de Kripke debemos reparar en algo importante: en nuestras consideraciones acerca de la semántica leibniziana nunca usamos ningún supuesto acerca de la estructura interna de los mundos posibles; sólo los usamos como «soportes» de distintos valores de verdad para la misma variable. Entonces podemos usar como «mundos» objetos cualesquiera. Lo importante es introducir todas las funciones f que se comporten de esta manera: f, aplicada a una variable proposicional y uno de los objetos elegidos, da como resultado un valor de verdad unívoco. Para asegurarnos de que se den todas las combinaciones entre mundos y valores de verdad de las infinitas letras proposicionales, debemos trabajar con un conjunto infinito fijo de «mundos» o con infinitos conjuntos diferentes de objetos. Eligiendo la última vía, podemos introducir la importante noción de *modelo de Kripke*. Un modelo de Kripke es una terna ordenada tal que su primer componente es un conjunto no vacío M, su segundo componente es una relación R entre miembros de M (*i.e.*, un subconjunto de M $\times$ M) y su tercer componente una función V que a cada par ordenado de una variable proposicional y un miembro de M le asigna un valor de verdad unívoco. Representaremos un modelo de Kripke arbitrario mediante la notación $<$M, R, V$>$. El par ordenado de los dos primeros componentes de un modelo de Kripke es un *marco* de Kripke. En las semánticas que estudiaremos a continuación, los modelos de Kripke cumplirán el rol de estructuras. Por razones similares a las que se adujeron en el caso de la semántica leibniziana, no se puede definir *verdad para un modelo* sino *verdad para un mundo en un modelo*. Las cláusulas recursivas de esta definición son como las de la semántica leibniziana, solo que reemplazando (iv) y (v) por (iv') y (v') y la idea de *estructura* por la de *modelo de Kripke*. Tenemos entonces:

(i) Si v es una variable proposicional, $<M, R, V>$ un modelo de Kripke y m un miembro de M, v es verdadera en m en ese modelo syss $V(v,m)$ = la verdad.

(ii) $\sim A$ es verdadera en un mundo m en un modelo de Kripke K syss A no es verdadera en m en K.

(iii) $(A \supset B)$ es verdadera en un mundo m en un modelo de Kripke K syss A no es verdadera en m en K o B es verdadera en m en K.

(iv) $\Box A$ es verdadera en m en un modelo de Kripke $<M, R, V>$ syss para todo mundo n de ese modelo, tal que mRn, A es verdadera en n en ese modelo.

(v) $\Diamond A$ es verdadera en un mundo m en un modelo de Kripke $<M, R, V>$ syss para algún mundo n de ese modelo tal que mRn, A es verdadera en n en ese modelo.

A es *falsa* en un mundo m en un modelo de Kripke K syss A no es verdadera en m en K.

A es *válida* en un modelo de Kripke K syss es verdadera en todo mundo m en K. Finalmente, llegamos a la definición de *validez de una fórmula en una clase de modelos de Kripke* (esta noción es importante porque cumplirá la misma función que en otras semánticas tiene la noción de *verdad para una familia de estructuras*). A es *válida en una clase de modelos de Kripke* syss es válida en todo modelo de la clase (*i.e.*, es verdadera para todo mundo en todo modelo de la clase).

Se puede construir una semántica adecuada para T de la siguiente manera. Las estructuras de nuestra semántica serán modelos de Kripke. Pero para que la semántica sea adecuada para T, deberemos escoger una familia determinada de tales modelos. Elegiremos la clase de modelos de Kripke cuya relación R (el segundo componente del modelo) es reflexiva, en el sentido de que todo mundo del modelo (todo elemento de M) tiene la relación R consigo mismo. Llamaremos *modelo de Kripke reflexivo* a los modelos de Kripke que satisfagan esta condición.

Puede observarse en ejemplos algo que se puede demostrar en general: los teoremas de T son válidos en la clase de todos los modelos reflexivos. Consideremos otra vez la fórmula (2) de la última serie de cinco antes analizada. Sabemos que es un teorema de T. Es obvio también que es válida en todo modelo de Kripke reflexivo: si el antecedente de (2) es verdadero en un mundo m de un modelo reflexivo K, p será verdadera en todo mundo accesible a m; pero m mismo es uno de esos mundos y se desprende entonces que p es verdadera en m. Como dijimos, y el lector puede comprobar como ejercicio en algunos ejemplos, lo mismo ocurre para todo teorema de T. Pero lo interesante es que la clase de modelos de Kripke elegida no es omni-acogedora. Si se toman teoremas de S4 o S5 que no son teoremas de T, resulta que no son válidos en la clase de modelos escogida. Considérese (4). Tómese un modelo de Kripke reflexivo donde hay tres mundos m_1, m_2 y m_3, tales que m_1Rm_2 y m_2Rm_3 pero m_1 no tiene la relación con m_3 (por supuesto, cada mundo tiene la relación consigo mismo; no se dan más conexiones que las enumera-

das). Supongamos que p es verdadero en los dos primeros mundos pero no en el tercero. $\Box p$ es verdadero en m_1, porque p es verdadero en todos los mundos accesibles a m_1 (él mismo y m_2); pero $\Box\Box p$ no es verdadero en el primer mundo, porque $\Box p$ no es verdadero en el segundo (ya que en un mundo accesible a él, m_3, no es verdadera p). Se desprende que el antecedente de (4) es verdadero pero su consecuente falso en m_1. (4) no es válida para la clase de modelos de Kripke reflexivos.

En general, puede probarse que A es teorema de T syss es válida en la clase de modelos de Kripke reflexivos. Esta clase permite entonces definir una semántica adecuada para T.

3. *Semánticas para S4 y S5*

Se puede construir una semántica adecuada para S4 si se toma una clase de modelos de Kripke cuya relación R satisface otras condiciones: es reflexiva, pero también transitiva. Llamemos modelos de Kripke reflexivo-transitivos a los que cumplen con esta doble condición. Estos modelos determinan un conjunto de fórmulas: las que son válidas respecto de esta clase. Puede observarse que el modelo que usamos antes para mostrar la no-validez de (4) ya no servirá para probar la no-validez de esa fórmula en la nueva semántica: en aquel ejemplo R no era transitiva. De hecho, (4) es válida en la semántica que acabamos de construir (el lector puede intentar probar por el absurdo esta afirmación). La nueva clase de modelos se comporta respecto de S4 como la clase de modelos reflexivos respecto de T: A es teorema de S4 syss es válida en la clase de los modelos de Kripke reflexivo-transitivos.

La semántica leibniziana que describimos antes era adecuada para S5, pero se crearía una discontinuidad inelegante si usáramos esa semántica para S5 y otras basadas en modelos de Kripke para T y S4. Los métodos de Kripke también son aplicables a S5. Esto es trivialmente verdadero si se piensa que la clase de modelos de Kripke cuya relación es universal (*i.e.,* que cualquier par ordenado de mundos del modelo pertenece a la relación) se comporta exactamente como las estructuras de nuestra semántica leibniziana. Otra manera de construir una clase de modelos de Kripke que proporciona una semántica adecuada para S5 es exigir que la relación del modelo sea una relación de equivalencia (reflexiva, simétrica y transitiva). Llamemos modelos de Kripke R-equivalentes a los que cumplen esta condición. Puede probarse respecto de ellos que A es válida en esa clase de modelos syss es teorema de S5.

En la literatura sobre lógica modal se dice que un sistema modal S es *correcto* respecto de una clase C de modelos de Kripke syss todo teorema de S es válido en la clase C. Se dice también que S es *completo* respecto de C syss toda fórmula de S que sea válida en C es teorema de S. Si un sistema modal S es correcto y completo respecto de una clase C de modelos de Kripke, se dice también que la clase C *caracteriza* el sistema S. En esta sección hemos introducido varias semánticas basadas en

clases de modelos de Kripke. Afirmar que una de ellas es adecuada para un sistema modal equivale a decir que la clase de modelos en que está basada caracteriza el sistema en cuestión. Para probar esto último se deben demostrar dos teoremas, uno de corrección y otro de completitud. Debido al bagaje de conocimientos que hemos presupuesto en el lector y el espacio disponible, no podemos probar aquí la adecuación de las semánticas que hemos introducido para T, S4 y S5. Los trabajos de Kripke que hemos mencionado en la sección II (en la parte sobre «la etapa semántica») proporcionan todos los elementos para construir las pruebas correspondientes. Una presentación didáctica de los teoremas relevantes y sus demostraciones puede encontrarse en Hughes y Cresswell (1968) y Hughes y Cresswell (1984). Un libro muy útil para profundizar en la semántica de la lógica modal es Chellas (1980).

V. INVESTIGACIONES ACTUALES SOBRE METALÓGICA MODAL

En las investigaciones recientes sobre lógica modal ha predominado una tendencia a estudiar las propiedades de sistemas modales de diferentes clases o las relaciones del lenguaje modal con ciertas estructuras. En líneas generales puede decirse que el interés teórico se ha desplazado hacia lo que describí hacia el final de la sección II como «la metalógica modal generalizada». Muchos de los trabajos caen en dos áreas que se denominan «teoría de la completitud» y «teoría de la correspondencia».

En la teoría de la completitud se clasifican los conjuntos de fórmulas modales (de un lenguaje modal especificado). En la presentación de Jansana (1990), se llama *lógica* a todo conjunto de fórmulas modales que contenga todas las tautologías y esté cerrado para las reglas de *modus ponens* y substitución (*i.e.,* si ciertas fórmulas pertenecen al conjunto, también pertenecen a él las fórmulas que pueden inferirse de las primeras mediante las reglas mencionadas —que son las reglas MP y RS de mi formulación de T). Hay varios grupos importantes de lógicas. El grupo más estudiado es el de las lógicas normales. Una lógica es normal si contiene todos los ejemplos de substitución del axioma 5 de T, en la formulación de la sección III, y está cerrado para la regla de necesidad que aparece en esa misma formulación. Las fórmulas que pertenecen a una lógica normal se dicen *teoremas* de esa lógica.

Recordemos que un marco de Kripke es el par ordenado de los dos primeros componentes de un modelo de Kripke. Un modelo de Kripke $<M, R, V>$ es un modelo del marco $<M, R>$. Una fórmula es válida respecto de un marco de Kripke cuando es válida respecto de todos los modelos de ese marco. La validez, corrección y completitud respecto de clases de marcos de Kripke se define de manera análoga a las nociones de validez, corrección y completitud respecto de clases de modelos de Kripke. Una clase de marcos de Kripke C caracteriza una lógica syss tal lógica es correcta y completa respecto de C. Algunas preguntas típicas

de la teoría de la completitud son: ¿Existe para cada lógica normal una clase de modelos que la caracterice? ¿Existe para cada lógica normal una clase de marcos que la caracterice? ¿Plantean el mismo problema las dos preguntas anteriores? ¿Existe alguna lógica normal caracterizada por la clase de los marcos irreflexivos (aquéllos en que la relación R es irreflexiva)? Estas y otras preguntas más técnicas han originado una buena cantidad de resultados interesantes de la teoría de la completitud.

La teoría de la correspondencia surgió a partir de la observación de ciertas conexiones entre sistemas modales y propiedades de la relación R de la clase de marcos que los caracterizaban. Por ejemplo, S4 es caracterizada por la clase de los marcos cuya relación R es reflexiva y transitiva. También fórmulas aisladas están conectadas con propiedades de la relación R de determinadas clases de marcos. Por ejemplo, el axioma 4 de nuestra formulación de T es válido en un marco de Kripke syss la relación R de tal marco es reflexiva. El axioma 6 de S4 (en nuestra presentación) es válido en un marco syss la relación de tal marco es transitiva. En las tres conexiones citadas, hay correspondencias entre sistemas o fórmulas modales y clases de marcos cada una de las cuales tiene como característica distintiva tener una relación R que cumple cierta condición expresable mediante una fórmula de primer orden (reflexividad y transitividad, reflexividad, o transitividad, en nuestros ejemplos). Estas correspondencias suscitan algunos interrogantes: ¿Toda fórmula modal está conectada con una propiedad de R expresable en la lógica de primer orden? En caso de que no sea así, ¿qué tipo de fórmulas modales determinan condiciones de primer orden? Estas y otras preguntas más técnicas originaron muchos hallazgos de la teoría de la correspondencia. Dos resultados parciales pueden ser de interés para el lector. Toda fórmula modal A en la que no hay operadores modales aplicados a fórmulas que ya tienen operadores modales, está conectada con una clase de marcos de Kripke determinada por una condición acerca de R expresable en la lógica de orden uno (la conexión es, naturalmente, que la fórmula es válida en un marco syss pertenece a la clase de marcos en cuestión). Hay fórmulas modales que no están conectadas con ninguna clase de marcos caracterizada por una propiedad de su relación R expresable en la lógica de orden uno. Un ejemplo es el llamado axioma de Löb:

$\Box(\Box p \rightarrow p) \rightarrow \Box p$

Se considera que la obra fundamental en la teoría de la completitud es Segerberg (1971). Goldblatt (1976) y Van Benthem (1982) contienen lo esencial de la teoría de la correspondencia. Van Benthem (1984) sintetiza muchos resultados. Jansana (1990) es una excelente introducción a la teoría de la completitud y la teoría de la correspondencia (en la que se tratan también otros temas de lo que he llamado «metalógica modal generalizada»; la obra no presupone ningún conocimiento de lógica modal, aunque quizás algún conocimiento de sistemas modales particulares pueda requerirse para que un lector se interese por los resultados más abstractos y generales de este libro).

APÉNDICE

Las limitaciones de espacio han hecho que me restringiera en este trabajo a lógica modal proposicional. La parte II de Hughes y Cresswell (1968) contiene una introducción básica a la lógica modal cuantificacional. Garson (1984) da información sobre investigaciones y resultados más recientes en este campo.

También se han omitido en este artículo enfoques algebraicos de la lógica modal, dado que en este volumen no se suministran elementos básicos de enfoques de tal tipo. Jansana (1990) también proporciona información sobre investigaciones en este terreno. Por último, han sido excluidas aquí las teorías que son *lógicas modales* sólo en un sentido amplio de la expresión (por ejemplo, lógicas deónticas, temporales, condicionales, etc.). Se puede encontrar información sobre ellas en otros artículos de este volumen y en Gabbay y Guenthner (1984).

BIBLIOGRAFÍA

Anderson, A. R. y Belnap, N. D. (1975), *Entailment: The Logic of Relevance and Necessity,* vol. I, Princeton University Press, Princeton.

Aristóteles (1984), *The Complete Works of Aristotle* (la versión de Oxford revisada por Barnes), vol. I, Princeton University Press, Princeton.

Aristóteles (1988), *Tratados de lógica (Organon),* dos volúmenes, v.e. Candel Sanmartín, Gredos, Madrid.

Becker, O. (1930), «Zur Logik der Modalitäten»: *Jahrbuch für Philosophie und Phänomenologische Forschung, 11,* 497-548.

Bull, R. y Segerberg, K. (1984), «Basic Modal Logic», en Gabbay y Guenthner, 1984.

Carnap, R. (1947), *Meaning and Necessity,* The University of Chicago Press, Chicago.

Carnap, R. (²1956a), *Meaning and Necessity,* The University of Chicago Press, Chicago-London.

Carnap, R. (1956b), «Meaning and Synonymy in Natural Languages», en Carnap (1956a).

Chellas, B. (1980), *Modal Logic,* CUP, Cambridge.

Church, A. (1943), «Reseña de Quine (1943)»: *Journal of Symbolic Logic, 8,* 45-47. También incluida en versión española en Simpson, 1973, con el título «Acerca del artículo de Quine, 'Notas sobre existencia y necesidad'».

Feys, R. (1937), «Les logiques nouvelles des modalités»: *Revue Néoscholastique de Philosophie, 40,* 517-553.

Forbes, G. (1985), *The Metaphysics of Modality,* Clarendon Press, Oxford.

Gabbay, D. y Guenthner, F. (compiladores) (1984), *Extensions of Classical Logic* (vol. II de Gabbay y Guenthner, *Handbook of Philosophical Logic*), D. Reidel Publishing Company, Dordrecht.

Garson, J. (1984), «Quantification in Modal Logic», en Gabbay y Guenthner, 1984.

Gödel, K. (1933), «Eine Interpretation des intuitionistischen Aussagenkalküls»: *Ergebnisse eines mathematischen Kolloquiums, 4,* 34-40.

Goldblatt, R. (1976), «Metamathematics of Modal Logic»: *Reports on Mathematical Logic, 6,* 41-77, *7,* 21-52.

Hintikka, J. (1961), «Modality and Quantification»: *Theoria, 27,* 119-128. También incluido en Hintikka, 1969, en una versión ampliada.

Hintikka, J. (1963), «The modes of modality»: *Acta Philosophica Fennica, 16,* 65-82.
Hintikka, J. (1969), *Models for modalities,* D. Reidel Publishing Company, Dordrecht.
Hughes, G. y Cresswell, M. (1968), *An Introduction to Modal Logic,* Methuen and Co. Ltd., Londres (conviene consultar la reimpresión 1972 o posteriores, porque en ellas se ha corregido una prueba de completitud defectuosa de la primera impresión). V. e., *Introducción a la lógica modal,* Tecnos, Madrid, 1973.
Hughes, G. y Cresswell, M. (1984), *A companion to modal logic,* Methuen, London-New York.
Jansana, R. (1990), *Una introducción a la lógica modal,* Tecnos, Madrid.
Kanger, S. (1957), *Probability in Logic,* Almqvist y Wiksell, Stockholm.
Kaplan, D. (1968), «Quantifying»: *Synthese, 19,* 178-214. V. e. en Simpson, 1973, bajo el título «Cuantificación, creencia y modalidad».
Kneale, W. y Kneale, M. (1962), *The development of Logic, OUP,* Oxford (v.e. *El desarrollo de la lógica,* Tecnos, Madrid, 1972).
Knuuttila, S. (comp.) (1988), *Modern Modalities,* Kluwer Academic Publishers, Dordrecht.
Knuuttila, S. (1993), *Modalities in Medieval Philosophy,* Routledge, Londres y Nueva York.
Kripke, S. (1959), «A completeness theorem in modal logic»: *Journal of Symbolic Logic, 24,* 1-14.
Kripke, S. (1963a), «Semantical Analysis of modal logic I»: *Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 9,* 67-96.
Kripke, S. (1963b), «Semantical considerations on modal logics»: *Acta Philosophica Fennica, 16,* 83-94.
Kripke, S. (1980), *Naming and Necessity,* Basil Blackwell, Oxford (v. e. M. Valdés, *El nombrar y la necesidad,* UNAM, México, 1985).
Lemmon, E. y Scott, D. (1977), *An Introduction to Modal Logic,* Blackwell, Oxford.
Lewis, C. I. (1912), «Implication and the algebra of Logic»: *Mind N. S., 21,* 522-531.
Lewis, C. I. (1918), *A Survey of Symbolic Logic,* University of California Press, Berkeley. (Hay una edición de Dover muy posterior que no contiene el capítulo sobre implicación estricta).
Lewis, C. I. y Langford, C. (1932), *Symbolic Logic,* Dover, New York.
MacColl, H. (1880), «Symbolical reasoning»: *Mind, 5,* 45-60.
McCall, S. (1963), *Aristotle's modal syllogisms,* North Holland Publishing Co., Amsterdam.
Mendelson, E. ([3]1987), *Introduction to Mathematical Logic,* Wadsworth, Inc., Belmont, California.
Orayen, R. (1989), *Lógica, significado y ontología,* UNAM, México.
Parry, W. (1939), «Modalities in the *Survey* system of strict implication»: *Journal of Symbolic Logic, 4,* 131-154.
Quesada, D. (1991), «Lógica clásica de primer orden», en este volumen.
Quine, W. (1943), «Notes on Existence and Necessity»: *The Journal of Philosophy, 40,* 113-127. V. e. en Simpson, 1973.
Quine, W. (1951), «Two dogmas of empiricism», en Quine, 1961a, y Quine, 1980. (Había aparecido en 1951 en *Philosophical Review,* pero la versión que se recoge en los libros citados tiene cambios de alguna importancia.)
Quine, W. (1960), *Word and Object,* MIT (v.e. M. Sacristán Luzón, *Palabra y objeto,* Labor, Barcelona, 1968).
Quine, W. ([2]1961a), *From a logical point of view,* Harvard University Press, Cambridge (los mayores cambios de esta edición respecto de la primera están en el artículo en que se tratan las críticas de Quine a la lógica modal). V. e. M. Sacristán Luzón, *Desde un punto de vista lógico,* Ariel, Barcelona, 1962.
Quine, W. (1961b), «Reference and modality», en Quine, 1961a.
Quine, W. (1980), *From a logical point of view,* Harvard University Press, Cambridge, reimpresión de Quine (1961a) pero con una ligera modificación en el ensayo en que se ocupa de lógica modal.

Sainsbury, M. (1991), *Logical Forms,* Blackwell, Oxford.
Segerberg, K. (1971), *An Essay in Classical Modal Logic,* Philosophical Studies, University of Uppsala, Uppsala.
Simpson, T. (comp.) (1973), *Semántica filosófica: problemas y discusiones,* Siglo XXI, Buenos Aires-Madrid.
Sobociński, B. (1953), «Note on a modal system of Feys-Von Wright»: *The Journal of Computing Systems, 1,* 171-178.
Tomás de Aquino (1991), *Suma contra los gentiles,* Porrúa, México.
Van Benthem, J. (1982), *Modal Logic and Classical Logic,* Bibliopolis, Napoli.
Van Benthem, J. (1984), «Correspondence Theory», en Gabbay y Guenthner, 1984.
Van Rijen, J. (1989), *Aspects of Aristotle's Logic of Modalities,* Kluwer Academic Publishers, Dordrecht.
Whitehead, A. y Russell, B. (1910), *Principia Mathematica,* CUP, Cambridge, (1.ª ed.). El material de los *Principia* que más se consulta se puede encontrar en la versión abreviada (basada en la segunda edición) *Principia Mathematica to *56,* CUP, 1967.
Wright, G. von (1951), *An Essay on Modal Logic,* North Holland Publishing Co., Amsterdam. V. e. A. Demarchi, *Ensayo sobre la lógica modal,* Rueda, Buenos Aires, 1970.

LÓGICAS MULTIVALENTES

Lorenzo Peña

I INTRODUCCIÓN HISTÓRICA

La idea central subyacente a la construcción de lógicas multivalentes es la de que hay un cierto campo fronterizo entre la verdad total y la completa falsedad. Esa idea no es ningún invento de algunos lógicos contemporáneos, sino que tiene hondas y remotas raíces en el pensamiento humano, y cabe alegar a su favor muchas consideraciones de muy diversa índole, desde las puramente filosóficas hasta las referidas a dificultades surgidas en no pocas disciplinas científicas por la pretensión de encasillar cada situación en uno de los dos polos, o «valores de verdad», de la lógica clásica.

Sin remontarnos a pensadores a quienes, como a Heráclito y a Platón, cabe fundadamente atribuir la propuesta de situaciones intermedias entre esos dos polos o extremos —en el caso de Platón con su tesis de grados de verdad o de realidad—, hay algún indicio a cuyo tenor no pareciera descaminado ver en Raimundo Lulio y en Nicolás de Cusa, entre otros, esbozos, todo lo rudimentarios que se quiera, de algo así como lógicas multivalentes. Sin embargo, fue uno de los fundadores de la lógica contemporánea, Charles S. Peirce, quien, junto con muchos otros logros, esbozó claramente, por vez primera, un sistema de lógica trivalente y además elaboró argumentos filosóficos convincentes a su favor. Sus apuntes al respecto recorren un largo lapso, mas en cualquier caso se sabe que en 1909 desarrolló esas ideas y alcanzó resultados rigurosos. Su plan de una matemática triádica o tricotómica concebía la inclusión del dominio limítrofe entre la afirmación y la negación «positivas» como un ensanchamiento más que como un debilitamiento de la lógica clásica (el principio de tercio excluso no había de venir omitido, pero sí reinterpretado de tal forma que no fuera enteramente verdadero). Peirce no publicó esos esbozos, desgraciadamente, y su obra no influyó en el ulterior [re]nacimiento de las lógicas multivalentes. (Sobre el aporte de Peirce, vid. Rescher, 1969,

4-5; ese mismo libro es la mejor fuente bibliográfica y de referencia para buena parte de las someras indicaciones de esta sección.)

El primer sistema estricto de lógica multivalente en ser dado a conocer en público fue el sistema trivalente del lógico polaco Jan Łukasiewicz en 1920 (ver Łukasiewicz, 1967). Durante los años 20 el propio Łukasiewicz y otros lógicos polacos desarrollaron ese sistema y fueron inventando otros con más de tres valores de verdad. Uno de esos lógicos, M. Wajsberg, brindó en 1932 una axiomatización completa para la lógica trivalente de Łukasiewicz: éste, por su parte, ya en 1930 expresó su preferencia filosófica por la lógica infinivalente. Siguieron en años sucesivos numerosos trabajos de estudio sintáctico y semántico de esos y otros sistemas multivalentes por diversos lógicos polacos, como J. Słupecki, Boleslaw Sobociński, St. Jaśkowski, etc.

Independientemente, el lógico norteamericano E. Post inventó en 1921 otro sistema diferente de lógica trivalente. Luego generalizó su tratamiento a m valores (para m finito). Desde el punto de vista lógico, suscita una dificultad el tratamiento de Post, y es que lo que él propuso fue una lógica, no de enunciados, sino de conjuntos de enunciados, por lo cual no resulta fácil entender sus sistemas como cálculos sentenciales. Pero dieron lugar a estudios algebraicos que luego se han revelado fructíferos. En ese orden del estudio algebraico han abundado cada vez más las contribuciones destacadas, entre las que cabe citar las de Gregor Moisil ya antes de la II guerra mundial (en Moisil, 1972), y luego Balbes & Dwinger (1974), Varlet (1975), Rasiowa (1974), el matemático portugués Antonio Monteiro y su colaborador y discípulo argentino —radicado durante un tiempo en el Brasil— R. Cignoli (1980).

Otro aporte muy original fue el de S.C. Kleene, cuyo sistema lógico trivalente, de 1938, presentaba rasgos que lo separaban, interesantemente, de los de Łukasiewicz. Igualmente original era el sistema trivalente del lógico ruso Bochvar (propuesto en 1939), que postulaba 3 valores, V, F e I, y que atribuía I a cada fórmula no atómica que tuviera entre sus componentes una fórmula con valor I. Ese sistema no tiene tautologías, pero recientemente Urquhart (1986) ha probado su interés desde el punto de vista de la teoría de pruebas.

Urquhart aborda ese estudio y el de otros cálculos multivalentes —incluido uno que él propone y que ha sido desarrollado en Méndez (199?)— utilizando un nuevo y más fecundo enfoque, que es el *matricial*, un cruce entre teoría de pruebas y lógica algebraica, y que consiste en dilucidar qué relaciones de consecuencia vienen determinadas por la asignación de ciertas álgebras como modelos para los cálculos lógicos que se trate de estudiar. Ese tratamiento matricial —dentro del cual se ubica el presente estudio también— ha sido desarrollado, entre otros, por Malinowski (1979), Rautenberg (1979)

y sobre todo el ya citado Urquhart.

No sin conexión con los desarrollos ya mencionados, tuvieron lugar otros que no pueden dejar de considerarse dentro del ámbito de las lógicas multivalentes: la axiomatización de la lógica intuicionista por Heyting en 1930 (desde el punto de vista algebraico ese cálculo se caracteriza por álgebras pseudocomplementadas, en el sentido indicado en el §2 de este trabajo) y más aún los cálculos propuestos por Kurt Gödel en 1932 (cuya contraparte algebraica son álgebras de Stone; ver §2, *infra*); sobre esos aportes, ver Rescher (1969). Otra exploración de las lógicas multivalentes se efectuó con vistas al tratamiento de anomalías en la física cuántica; fue iniciada en 1937 por P. Détouches-Février y desarrollada por Reichenbach en 1944 (ver Haack, 1974, 148 ss, 172-4).

Una auténtica explosión de estudios y de aplicaciones de lógicas multivalentes ha tenido lugar desde que en 1965 el trabajo pionero del ingeniero electrónico californiano Lofti Zadeh (ver Zadeh y ot. (1975)) inauguró el tratamiento de las *lógicas de lo difuso*, y de las *teorías de conjuntos difusos*. La idea central (que ya antes había sido propuesta, entre otros por Rescher) es tomar como función característica de un conjunto una que tome sus valores o imágenes en un conjunto de más de dos valores de verdad —preferiblemente en un dominio de infinitos valores. Aunque enfoques de ese género no han suscitado ni mucho menos unanimidad y siguen siendo ásperamente controvertidos, numerosísimos científicos de las más variadas disciplinas han abrazado con ardor ese tipo de tratamientos, habiéndoles encontrado, o creído encontrar, múltiples aplicaciones en sus respectivos campos. Más que nada descuella en esa porfía la informática, donde, curiosamente, el binarismo que parecía subyacente de manera definitiva se ha visto así contrarrestado o acaso completado por los tratamientos multivalentes. El hecho es que quienes más han contribuido a propagar el uso y cultivo de las viejas y nuevas lógicas multivalentes han sido los ingenieros electrónicos. Desde esa perspectiva han surgido un sinfín de nuevos tratamientos algebraicos, p.ej. (destaca aquí el grupo barcelonés de Enrique Trillas, L. Valverde y otros; ver, a título de ejemplo no más, Trillas & Valverde (1982)). No sin parentesco con esa línea de estudios están los nuevos tratamientos de inteligencia artificial y temas conexos utilizando lógicas paraconsistentes multivalentes, como las *lógicas anotadas* (en las que una fórmula «dice» de algún modo qué valor o grado de verdad posee: ver da Costa, Subrahmanian & Vago (1989); ello guarda afinidad con lo esbozado en el §2, hacia el final). Entre las lógicas multivalentes que son a la vez paraconsistentes cabe asimismo mencionar una lógica trivalente que ha sido separada e independientemente descubierta e investigada por varios autores; entre ellos el autor de estas páginas, por un lado, y por otro

da Costa e Ítala d'Ottaviano, conjuntamente; ésta última la ha estudiado a fondo en su tesis d'Ottaviano (1982).

Otra área donde ha prosperado la lógica multivalente es la del tratamiento semántico de las lógicas relevantes, desarrolladas desde los años 70; ver Anderson & Belnap (1975). La relación entre lógicas multivalentes y lógicas de la relevancia ha sido investigada, p.ej., por Urquhart, en su trabajo ya citado, y por Sylvan & Urbas (1989). Alguna de las lógicas estudiadas en este último trabajo guardan estrechísimo parentesco con las que más centrarán nuestra atención en la segunda mitad o así del §2. (Sobre ese parentesco, volveré justamente al final del §2.)

Hasta que se empezó a trabajar en teorías de conjuntos difusos prevalecía en el estudio de lógicas multivalentes la preferencia por lógicas con un número finito de valores. Pero para su aplicación a la teoría de conjuntos, se han visto las ventajas de la infinivalencia. Desgraciadamente, sin embargo, resultaba muy difícil dar un tratamiento axiomático adecuado a cálculos cuantificacionales infinivalentes (la extensibilidad cuantificacional de las lógicas multivalentes en general se venía investigando desde hacía tiempo, sobresaliendo el aporte de Rosser & Turquette (1952), c. V, 62 ss; mas la prueba de la inaxiomatizabilidad del sistema $Ł_{\aleph}Q$ de lógica cuantificacional basado en el cálculo infinivalente de Łukasiewicz fue proporcionada por B. Scarpellini en 1962; ver la referencia en Urquhart, 1986, 99). Ello ha alejado a una parte de los estudiosos y cultivadores de esas teorías de conjuntos del tratamiento axiomático. Recientemente se ha puesto en pie una nueva familia de lógicas infinivalentes (y paraconsistentes) en la cual se obtiene la extendibilidad axiomática al cálculo cuantificacional, y además se prueba que uno al menos de los sistemas de esa familia, **A**, es, para cada lógica *L* caracterizada por m valores de verdad, una extensión cuasi-conservativa de la misma, en el sentido de que hay en el sistema algún functor de afirmación generalizada, ☼, tal que, para cualquier fórmula ⌜p⌝, ⌜☼p⌝ es un teorema de **A** sys (si, y sólo si) ⌜p⌝ es un teorema de *L* (ver Peña, 1991, 139 ss); como casos particulares se tiene una extensión del mismo resultado para la lógica $G_{\aleph}$ (el sistema infinivalente de Gödel) y otro incluso más fuerte, y es que **A** es una extensión conservativa de la lógica bivalente clásica (naturalmente, sólo para cierta traducción de la negación clásica).

Si la investigación de lógicas multivalentes ha suscitado entusiasmo, no han faltado sus detractores, quienes han tendido a ver en esos cálculos invenciones artificiales y sin base «intuitiva», o incluso exentos de interés matemático. Como qué tautologías se den en un sistema multivalente y también qué relación de consecuencia haya en él dependen de qué valores sean tomados como *designados* (o verdaderos), unos cuantos autores han concluido

que se trata de algo meramente arbitrario, y por ende que todo el tratamiento ofrecido por tales lógicas es un juego. Cae naturalmente fuera del ámbito del presente trabajo discutir las motivaciones filosóficas, pero el hecho es que éstas existen, y a favor de ellas abonan muchos argumentos propuestos por diversos autores. Sobre ese y otros puntos que han de quedar fuera del presente estudio, ver Peña (1994).

Hoy se suele estar de acuerdo —más allá de tantísimas discrepancias en tantas cosas— al menos en esto: que el tratamiento de lógicas llamadas multivalentes forma parte del estudio algebraico de la lógica. En verdad hay una prueba trivial (generalización de un resultado célebre de Lindenbaum) a cuyo tenor cualquier sistema tiene una matriz característica multivalente: basta con tomar como álgebra una cuyo portador sea el conjunto de las fórmulas, y cuyos elementos designados sean los teoremas. En ese sentido no hay lógica que no sea multivalente. Ese resultado no banaliza el estudio de las lógicas multivalentes porque, precisamente, el tratamiento algebraico permite ver qué reducibilidades ulteriores se dan (por vía de congruencias, noción que será explicada en el §2).

II LAS LÓGICAS MULTIVALENTES COMO LÓGICAS ALGEBRAICAS

Las convenciones notacionales usadas aquí son, esencialmente, las de Church: un punto indica un paréntesis de abrir cuyo correspondiente paréntesis de cerrar estaría tan a la derecha como quepa; las restantes ambigüedades se disipan asociando hacia la izquierda

Por *álgebra universal* cabe entender un conjunto dotado de ciertas operaciones, siendo una operación una función n-aria, para $n \geq 0$ (n puede ser infinito, pero aquí excluiremos tal posibilidad). E.d., una operación n-aria, ¶, definida sobre un conjunto A, es algo tal que, para cualesquiera n miembros de A, a^1, a^2, ..., a^n, ¶(a^1...a^n) $\in$ A: lo cual significa que para a^1,...,a^n $\in$ A hay un solo miembro de A que es ¶(a^1,...,a^n). Cuando se trate de una operación binaria, en vez de ⌜¶(a^1,a^2)⌝ escribimos ⌜a^1¶a^2⌝ . Generalmente un álgebra se representa como una secuencia $<$A, $\Gamma$$>$, donde Γ es una secuencia de operaciones, o bien, alternativamente, como una secuencia $<$A, ¶^{1},...,¶^{j}$>$, donde ¶^{1},...,¶^{j} son operaciones ordenadas por su ariedad, o sea por el número de sus argumentos. El conjunto A será el *portador* de dicha álgebra. A veces, por comodidad, se llama al álgebra igual que a su portador.

Una *matriz* es un un trío $<$A, *D*, $\Delta$$>$, donde $<$A, $\Delta$$>$ es un álgebra —siendo A, Δ, según se acaba de indicar—, al paso que *D* es un subconjunto

de A, subconjunto que viene llamado el conjunto de elementos *designados*.

El procedimiento general para establecer la correspondencia entre un cálculo lógico y una matriz es éste. Dos álgebras se llaman *similares* [entre sí], o de de similaridad igual, si son, respectivamente, $<A, \Gamma>$ y $<B, \Delta>$, y $\Gamma = <\gamma_1, \gamma_2, ..., \gamma_n>$ mientras que $\Delta = <\delta_1, \delta_2, ..., \delta_n>$ y, para cada índice, i ($1 \leq i \leq n$) γ_i es una operación de la misma ariedad que δ_i, o sea son operaciones con el mismo número de argumentos. Un *morfismo* de un álgebra $<A, \Delta>$ en otra a ella similar $<B, \Gamma>$ es una función *f* tal que, para cada índice i, si δ_i es una operación m-aria, entonces para cualesquiera m elementos de A, a_i, ..., a_m, $f(\delta_i(a_1,...,a_m)) = \delta_i(fa_1,...,fa_m)$. Si el morfismo m es una *inyección* (mx=mz sólo si x=z), será llamado un *monomorfismo*; si es una *sobreyección* (para cada b∈B hay un a∈A tal que ma=b), será un *epimorfismo*; si es ambas cosas a la vez, un isomorfismo. Un morfismo de un álgebra en sí misma es un *endomorfismo*; un endomorfismo isomórfico es un automorfismo.

Un cálculo lógico [sentencial] viene definido como sigue. Un lenguaje sentencial es un álgebra de índole particular, a saber: una en la que el portador es un conjunto de fórmulas y las operaciones son simplemente las operaciones n-arias (para $n \geq 0$) que envían a n fórmulas, tomadas como argumentos, sobre la fórmula resultante de unir a las dadas mediante un functor n-ádico determinado; pueden ser, p.ej., las de negación, disyunción y conyunción. Si $<A, \Gamma>$ es un lenguaje sentencial, $<A, \Gamma, \Re>$ es un *cálculo* [sentencial] sys $\Re$ es una operación de consecuencia en A, donde una *operación de consecuencia* viene definida como una función ϕ que toma como argumentos subconjuntos de A y que cumple estas condiciones: 1ª) $\phi\phi X = \phi X \supseteq X$; 2ª) $\phi Z \not\subseteq \phi X$ sólo si $Z \not\subseteq X$; 3ª) ningún endomorfismo m es tal que $m\phi X \not\subseteq \phi m X$. Llamamos *regla de inferencia* de un cálculo cuyo portador (conjunto de fórmulas) sea A a una relación R entre dos subconjuntos de A que cumpla estas tres condiciones: 1ª) R se mantiene para cualquier endomorfismo —o sea: si X guarda R con Z, y m es un endomorfismo, siendo $m(X) = \{q: \exists r \in X(mr=q)\}$, entonces m(X) guarda R con m(Z); 2ª) si $\Re$ es la operación de consecuencia definitoria del cálculo C, entonces X guarda R con {z} sólo si $z \in \Re X$; 3ª) un conjunto X guarda con otro Z la relación R sólo si para cierto z Z={z}; a efectos prácticos podemos representar el que X guarde la relación R con $\{ \ulcorner q \urcorner \}$ como el que se dé esa relación R entre X y la fórmula $\ulcorner q \urcorner$. Que $\ulcorner q \urcorner \in \Re X$ lo escribimos: $X \vdash q$. Si $X = \{ \ulcorner p^1 \urcorner, ..., \ulcorner p^n \urcorner \}$, podemos expresar lo mismo así: $p^1,...,p^n \vdash q$. Decimos que una operación σ es la *extensión ancestral* de una familia de relaciones $\{R_i\}_{i \in I}$ sys σX es el menor superconjunto de X cerrado con respecto

a cada relación R_i. Una operación de consecuencia $\mathfrak{R}$ será llamada *regular* sys hay un número finito de reglas de inferencia, $R_1,...,R_m$, tales que $\mathfrak{R}$ es la extensión ancestral de $\{R_1,...,R_m\}$. En tal caso diremos que $\mathfrak{R}$ es la operación de consecuencia engendrada por $R_1,...R_n$. Sólo nos interesaremos aquí por operaciones de consecuencia regulares, lo cual nos permitirá, de hecho, pensar, más que en la operación en sí, en las reglas de inferencia que la engendran.

Aquellos elementos del portador de un cálculo con los cuales el subconjunto *vacío* de fórmulas, Ø, guarde alguna regla de inferencia de ese cálculo son sus *axiomas*. Los *teoremas* de ese cálculo son los miembros del menor superconjunto del conjunto de sus axiomas cerrado para la operación de consecuencia $\mathfrak{R}$. A un cálculo sentencial lo llamaremos también una *lógica*. (Nótese que está lejos de ser baladí la estipulación de este párrafo —aparentemente sólo definicional—: de hecho el concebir así a los axiomas equivale a adoptar automáticamente la concepción clásica de la operación de consecuencia, a tenor de la cual, para cualquier conjunto de fórmulas X y cualquier teorema del cálculo sentencial considerado, ⌜p⌝, se tendrá que ⌜p⌝ $\in \mathfrak{R}$X —donde $\mathfrak{R}$ es la operación de consecuencia definitoria de ese cálculo sentencial).

Llamaremos *valuaciones* a los morfismos de un cálculo en una matriz, y *sustituciones* a los endomorfismos de un cálculo. Para que haya una valuación de un cálculo dado, *C*, en una matriz dada, **M**, a cada functor n-ádico, ♪, de *C* le habrá de corresponder una operación n-aria, ♫, en **M** tal que, para cualquier valuación *v* del lenguaje en el que se formule *C* en **M** y cualesquiera fórmulas ⌜p^1⌝, ..., ⌜p^n⌝ del lenguaje de dicho cálculo, se tendrá: *v*(⌜♪(p^1,...,p^n)⌝) = ♫(*v*(⌜p^1⌝),...,*v*(⌜p^n⌝)). Por comodidad —y no prestándose ello a ningún equívoco— cabe escribir igual el signo '♪' de *C* y el que nombra a la operación ♫ de **M**.

Diremos que una matriz **A** = $<A,D,\{\delta_1,...,\delta_m\}>$ es un *modelo* de un cálculo $C = <\mathbf{C},\{\gamma_1,...,\gamma_m\},\mathfrak{R}>$ sys cada valuación de **C** en **A**, *v*, es tal que —definiendo *v*(X) como {z: ∃u(*v*u = z)}, para un X ⊆ C— ⌜p⌝ ∈ **T** [siendo **T** el conjunto de teoremas de **C**] sólo si *v*(⌜p⌝) ∈ *D*. Llamaremos a un modelo *recio* si cumple esta condición adicional: cada valuación *v* y cada fórmula ⌜p⌝ son tales que *v*(⌜p⌝) ∈ *D* si *v*(X) ⊆ *D* y ⌜p⌝ ∈ $\mathfrak{R}$X. Para que una matriz sea un modelo de un cálculo basta con que las valuaciones les den a los teoremas, como imágenes suyas, elementos designados; para que sea recio es, además, menester que la operación de consecuencia del cálculo sea volcada por *cada* valuación en una relación, dentro de la matriz dada como modelo, que preserve el estatuto de designación. En adelante tan sólo

nos interesarán los modelos recios, por lo cual omitiremos el adjetivo. (Un modelo puede no ser característico —esta noción se va a definir unas pocas líneas más abajo—; y lo propio le sucede a una clase entera de modelos. Mas cuando se *define* un cálculo con relación a un modelo o clase de modelos, éste o éstos son entonces, por definición, característicos.)

En general un cálculo **C** del lenguaje $\mathscr{L}$ puede venir [semánticamente] caracterizado (definido) con relación a una clase **M** de matrices así: la operación [regular] de consecuencia $\mathfrak{R}$ vendrá definida así: $\forall A \in \mathbf{M}$, dado un conjunto de fórmulas cualquiera, $X \subseteq \mathscr{L}$, $\mathfrak{R}X = \{x \in \mathscr{L}: \forall h \in \mathbf{Val}(\mathscr{L},A)$: $hx \in D$ o $h(X) \not\subseteq D\}$, donde $\mathbf{Val}(\mathscr{L},A)$ es el conjunto de valuaciones de $\mathscr{L}$ en A. (Como caso particular, los teoremas de **C** serán las fórmulas ⌜p⌝ de su lenguaje, $\mathscr{L}$, que sean *válidas* respecto a **M**, o sea tales que $\forall A \in \mathbf{M}$ $\forall v \in \mathbf{Val}(\mathscr{L},A)$ ocurra que v(⌜p⌝)$\in D$, siendo D el conjunto de elementos designados de A.) Para una $\mathfrak{R}$ definida semánticamente con respecto a la clase **M** de matrices, el que ⌜p⌝ $\in \mathfrak{R}X$ viene expresado así: $X \models_{\mathbf{M}} p$.

Cuando se cumplen las condiciones recién indicadas, decimos que esa clase **M** de matrices es *característica* del cálculo en cuestión. Pero también es característica de un cálculo una clase de matrices, aunque el cálculo no venga definido así, siempre que *pueda serlo* (o sea siempre que otra definición tenga la misma extensión). Cuando una cierta clase unitaria de matrices sea característica de un cálculo sentencial dado, se dirá que la única matriz perteneciente a esa clase unitaria es *característica* de dicho cálculo. Si el portador de la matriz tiene exactamente n miembros se llama n-valente a dicho cálculo. (Dicho de otro modo: la matriz M es característica del cálculo C sys se cumple en general esta condición: $X \models_{\{M\}} p$ sys ⌜p⌝ $\in \mathfrak{R}X$, donde $\mathfrak{R}$ sea la operación de consecuencia definitoria de C.)

De hecho no cualquier álgebra ofrece interés como modelo para los propósitos recién indicados. Prácticamente se consideran pertinentes aquellas álgebras que son *retículos*, o sea álgebras $\langle A, +, \bullet \rangle$ donde para cualesquiera x, z, u $\in$ A se tienen estas ecuaciones:
idempotencia: $x+x = x = x\bullet x$; *conmutatividad*: $x+z = z+x$; $x\bullet z = z\bullet x$
asociatividad: $x+z+u = x+(z+u)$; $x\bullet z\bullet u = x\bullet(z\bullet u)$;
absorción: $x+z\bullet x = x = x\bullet z+x$

Es más, para nuestros propósitos, sólo se aceptan como modelos retículos distributivos, o sea que cumplan la condición de distributividad, a saber: $x+(u\bullet z) = (x+u)\bullet(x+z)$ así como $x\bullet(u+z) = (x\bullet u)+(x+z)$. La operación + es llamada la *junción* y corresponderá a la disyunción, mientras que • viene llamado el *cruce* y corresponde a la conyunción. En un retículo el que $x = x\bullet z$ se expresa así también: $x \leq z$.

Si la correspondencia entre esos signos lógicos y esas operaciones algebraicas suele sustraerse a la controversia, no sucede lo propio con respecto a la negación. Es bastante común, sin embargo, el postular que ésta coresponda a una operación algebraica unaria, $\sim$, tal que se cumplan tres ecuaciones adicionales: *De Morgan* (o sea $\sim(x \bullet z) = \sim x + \sim z$, así como $\sim(x+z) = \sim x \bullet \sim z$) e *involutividad*: $\sim(\sim x) = x$. No obstante, para algunos cálculos, como veremos, se atenúan estas dos condiciones o se reemplazan por otras menos estrictas.

Un álgebra con las operaciones $\bullet$, $+$ y $\sim$ que cumplan esas condiciones será llamada un *álgebra de De Morgan*. Supongamos ahora un álgebra de De Morgan que reúna esta condición adicional: hay en ella dos elementos, 1 y 0, tales que, en general, $x \bullet 1 = x$ mientras que $x \bullet 0 = 0$, $\sim 0 = 1$; un álgebra tal será un *álgebra cuasibooleana*. De entre las álgebras cuasibooleanas se llaman *álgebras de Kleene* a las que cumplen en general esta condición: $z + \sim z \geq x \bullet \sim x$.

Un retículo distributivo con 0 y 1 se llama *pseudocomplementado* si en él se da una operación unaria, $\neg$, tal que para cada x se tiene que $\neg x$ es el mayor elemento disjunto de x (o sea uno tal que para cualquier z $z \bullet x = 0$ sys $\neg x \geq z$). Un retículo pseudocomplementado es un álgebra de Stone sys cumple esta condición: para todo x, $\neg x + \neg\neg x = 1$. (En un retículo pseudocomplementado —omítense aquí las pruebas— valen estas ecuaciones: $x \bullet \neg x = \neg x \bullet \neg\neg x$; $\neg\neg x \geq x$; $x \bullet z > 0$ sys $\neg\neg x \bullet z > 0$; $\neg x > \neg z$ sólo si $z > x$; $\neg(x+z) = \neg x \bullet \neg z$; $\neg\neg x \bullet \neg\neg z = \neg\neg(x \bullet z)$; $\neg\neg(\neg\neg x + \neg\neg z) = \neg\neg(x+z)$; $\neg(\neg x + x) = 0$.)

Un álgebra de Kleene es un *álgebra booleana* si cumple esta condición: $x \bullet \sim x = 0$ y $x + \sim x = 1$. También cabe definir a las álgebras booleanas como retículos pseudocomplementados en los que para cada x: $x + \neg x = 1$; y como álgebras de Stone en las cuales $x = \neg\neg x$

Para nuestro propósito requeriremos que *D* —el conjunto de elementos designados— sea un *filtro propio*, o sea un subconjunto propio del portador del álgebra tal que, en general, se cumplan estas dos condiciones: 1ª) si $x \bullet z = x$, siendo $x \in D$, entonces $z \in D$; y 2ª) si $x, z \in D$, $x \bullet z \in D$.

Sea **B** la clase de todas las matrices booleanas. Demuéstrase este resultado: sea *C* un cálculo semánticamente definido como uno que tenga por axiomas a las fórmulas de cierto lenguaje (con conyunción, disyunción y negación) válidas en cualquier matriz de **B** y cuya operación de consecuencia semánticamente definida (del modo más arriba indicado) sea $\models_{\mathbf{M}}$; ese cálculo es idéntico a otro definido igual pero en el cual, en vez de **M**, se tome el álgebra con sólo dos elementos, 0 y 1, siendo $D = \{1\}$. El cálculo caracteri-

zado por las álgebras booleanas es **LC** (la lógica clásica). Por ello precisamente se da en llamar a **LC** la *lógica bivalente*.

Una lógica multivalente será un cálculo sentencial semánticamente definido cuyos modelos sean ciertas matrices no booleanas (aunque esto ha de entenderse, en general, como que no todas las matrices de la clase en cuestión serán booleanas). Un ejemplo será $K_{\aleph}$, a saber: el cálculo sentencial en el que son teoremáticas sólo todas las fórmulas válidas respecto a cualquier *matriz de Kleene*, donde una matriz de Kleene viene definida como un álgebra de Kleene con un conjunto cualquiera de elementos designados que sea un filtro propio y cuyas reglas de inferencia sean las preservadoras de la designación en esa clase de álgebras.

Pasemos ahora de $K_{\aleph}$ a otras lógicas construibles como extensiones de $K_{\aleph}$. En primer lugar, puede extenderse el cúmulo de reglas de inferencia del siguiente modo. Formamos el sistema K_{ω}, a saber la lógica semánticamente definida como teniendo sus modelos en una clase unitaria de matrices, $\{K\}$, donde $K = <\mathbf{R}, \{1\}, \Delta>$ siendo $\mathbf{R} = [0,1]$ (o sea el intervalo de los números reales r tales que $1 \geq r \geq 0$), y siendo $\Delta = \{+, \bullet, \sim\}$, donde $x \bullet z = min(x,z)$, mientras que $x+z = max\,(x,z)$ (el signo '+' no hace aquí las veces de la adición); la operación unaria $\sim$ viene definida así: $\sim 1 = 0$; $\sim 0 = 1$; para $1 > x > 0$ el logaritmo en base 2 de $\sim x$ es igual al logaritmo en base x de 2. Como signos del cálculo lógico usemos 'N' en vez de '$\sim$', '$\vee$' en vez de '+', y '$\wedge$' en vez de '$\bullet$'. Son teoremas de K_{ω} los mismos que de $K_{\aleph}$, pero en $K_{\aleph}$ no están ciertas reglas de inferencia que sí están en cambio en K_{ω}, como ésta (la regla de Cornubia, [mal]llamada de Escoto): p, Np $\models$ q. Tomemos ahora, en vez de K, A, definida como $<\mathbf{R}, D, \Delta>$, con Δ igual y $D =]0,1]$, o sea el conjunto de reales r tales que $1 \geq r > 0$. Al cálculo sentencial semánticamente definido como tomando sus modelos en $\{A\}$ lo llamamos lógica $\mathbf{A}_{\omega}$. Hay una serie de esquemas que son teoremáticos en $\mathbf{A}_{\omega}$ sin serlo en $K_{\aleph}$, como éstos: $\ulcorner N(p \wedge Np) \urcorner$, $\ulcorner p \vee Np \urcorner$ —respectivamente no contradicción y tercio excluso. Por otra parte, la regla de Cornubia no está en el nuevo sistema; éste es, pues, una lógica *paraconsistente*.

¿Qué pasa si, en vez de [0,1], tomamos $\{0, ½, 1\}$, siendo $D = \{½, 1\}$? Es obvio, por la definición dada, que $N½ = ½$. Llamemos $\mathbf{A}_3$ a la lógica semánticamente caracterizada con respecto a la clase unitaria cuyo único miembro es esta matriz de tres elementos. Salta a la vista que $\mathbf{A}_3 = \mathbf{A}_{\omega}$. Es más, cabe demostrar que se trata de la lógica semánticamente definible con respecto a la clase de matrices que son álgebras de Kleene en las cuales para todo x $x + \sim x \in D$.

Un problema que surge con las lógicas multivalentes hasta aquí con-

sideradas es que no tenemos en ellas ningún functor que exprese una relación condicional o de implicación. Ese problema no surge en **LC**, porque en ella podemos definir ⌜p⊃q⌝ como ⌜∼p∨q⌝, y similarmente con las operaciones booleanas correspondientes, obteniéndose los dos rasgos apetecibles para un condicional, a saber: (r°1°) el functor '⊃' así definido preserva la designación, e.d. posee la condición del *modus ponens* (para cualquier valuación *v*, si *v*(⌜p⊃q⌝) y *v*(⌜p⌝) son designados, también lo es *v*(⌜q⌝); y (r°2°) tiene la propiedad de la deducción: si p^1, ..., p^n, r ⊨ q, entonces $p^1,\ldots,p^n$ ⊨ r⊃q.

Hay dos procedimientos comunes para introducir functores condicionales en lógicas multivalentes. Uno consiste en introducirlos como primitivos. El otro estriba en introducir primero una negación fuerte, y luego definir, por medio de ella, el condicional, igual que se hace en **LC**. Voy a centrarme aquí en este segundo procedimiento.

Si $<A,\Delta>$ es un álgebra de Stone, $<A, D, \Delta>$ será una *matriz de Cragg*, donde *D* es el conjunto de miembros *densos* de *A*, o sea de aquellos elementos x tales que $\neg x=0$. Una operación que nos sirve entonces para definir el condicional es la que define $x \supset z$ como $\neg x \vee z$.

Llamemos *lógica pétrea* al cálculo sentencial cuyo vocabulario abarca signos sentenciales y las constantes de conyunción ('∧'), disyunción ('∨'), y negación fuerte ('¬), cuyos teoremas son las fórmulas válidas en cualquier matriz de Cragg y cuya única regla de inferencia es el *modus ponens* (p∨¬q, q ⊢ p). Definiendo en una lógica pétrea el condicional del modo clásico (⌜p⊃q⌝ abrevia a ⌜¬p∨q⌝), ese functor posee los dos deseados rasgos ya enumerados, (r°1°) y (r°2°). Es más: la lógica pétrea es idéntica a **LC**, a pesar de la diversa definición semántica de ambas. Un ejemplo de matriz de Cragg es ℭ, a saber: la que tiene como portador el intervalo [0,1], siendo • y + como vinieron definidos para K_ω, y siendo $\neg x = 0$ si $x>0$, $\neg 0=1$. Aunque esta álgebra no es booleana, es también un modelo de **LC**: una fórmula es un teorema del cálculo sentencial clásico sys es válida con respecto a la matriz ℭ. Como se ve, faltan a ℭ para ser booleana dos condiciones: en general no se tiene $\neg\neg x = x$; ni $x+\neg x = 1$.

El vínculo entre ℭ y un álgebra de Boole puede hacerse más claro como sigue. Llamamos *congruencia* en un álgebra a una relación [diádica] Θ que tenga, para toda operación n-aria, ¶, la *propiedad de sustitución*, a saber: si $x^1\Theta z^1$, ..., $x^n\Theta z^n$, entonces ¶$(x^1,\ldots x^n)$ Θ ¶$(z^1,\ldots,z^n)$. Una congruencia es una relación de equivalencia. Si Θ es una congruencia de un álgebra *A* cuyo portador es J, podemos obtener el álgebra cociente de *A* por Θ, a saber una cuyo portador es el conjunto de las clases de equivalencia $[x]_\Theta$ (o [x] a secas,

si el contexto desambigúa), siendo x un elemento cualquiera de J y teniéndose que en general z ∈ [x] sys xΘz. En el álgebra ℭ hay una congruencia Θ tal que xΘz si x>0<z o x=0=z. El álgebra cociente de ℭ por Θ es un álgebra booleana: es el álgebra cuyo portador es {0,1}. La idea principal en el paso de ℭ a esa álgebra booleana es que en ésta se toman como si fueran indiscriminables o indiscernibles todos los infinitos elementos densos (designados).

Sea **M** = <**M**,*D*,Δ> una matriz de Cragg, y para los mismos **M** y *D* sea <**M**,*D*,Φ> una matriz de Kleene. Sea Ξ = Δ∪Φ. Entonces <**M**,*D*,Ξ> será una **ξ-matriz**. Las ξ-matrices son los modelos que caracterizan a la lógica $\mathbf{A}_\xi$. O sea, $\mathbf{A}_\xi$ es aquel cálculo sentencial cuyos teoremas son las fórmulas, de un lenguaje sentencial dado, que vienen enviadas por cualquier valuación sobre elementos designados de una ξ-matriz, y cuyas reglas de inferencia son las que preservan la designación. (Por ser **M** una matriz de Cragg, *D* será el conjunto de todos los elementos densos.) Nótese que en una ξ-matriz ya no puede establecerse una congruencia como la considerada en el párrafo anterior. Sea Ξ la ξ-matriz cuyo portador es [0,1]. Aunque en esta ξ-matriz 1Θ½, no sucede empero que N1 Θ N½, puesto que N1=0, N½=½, y sin embargo no ocurre que 0Θ½ (hay infinidad de elementos z congruentes con 1 y tales que Nz no es congruente con N1; el caso del elemento ½ se aduce sólo por ser un caso «extremo»). La lógica $\mathbf{A}_\xi$ es una lógica paraconsistente (con respecto a la negación simple 'N') y también multivalente; no es paraconsistente con respecto a la negación fuerte, ¬. En la lógica $\mathbf{A}_\xi$ se tiene la regla de inferencia: p, ¬p ⊨ q; pero no: p, Np ⊨ q. Sin embargo, los principios de no contradicción y de tercio excluso son válidos para ambas negaciones: son teoremáticos todos los esquemas: ⌜p∨Np⌝, ⌜p∨¬p⌝, ⌜N(p∧Np)⌝, ⌜¬(p∧¬p)⌝. Definiendo ⌜p≡q⌝ como ⌜p⊃q∧.q⊃p⌝, se tendrá ⌜p≡¬¬p⌝, versión atenuada de la involutividad. (También se tiene, claro ⌜p≡NNp⌝.) Asimismo valen en la lógica $\mathbf{A}_\xi$ los cuatro principios condicionales de abducción, a saber: ⌜p⊃Np⊃Np⌝, ⌜p⊃¬p⊃¬p⌝, ⌜Np⊃p⊃p⌝, ⌜¬p⊃p⊃p⌝. Aun siendo paraconsistente, $\mathbf{A}_\xi$ es una lógica de talante muy conservador: su negación simple posee la mayor parte de los rasgos de la negación clásica, al paso que, gracias a su negación fuerte, $\mathbf{A}_\xi$ es una extensión conservativa recia de **LC**, en el sentido técnico usual (una fórmula de $\mathbf{A}_\xi$ que sólo contenga vocabulario clásico es teoremática en $\mathbf{A}_\xi$ sys también lo es en **LC**; y cada regla de inferencia clásica vale también en $\mathbf{A}_\xi$).

Demostrablemente la lógica $\mathbf{A}_\xi$ se caracteriza por la clase unitaria que sólo abarca a una matriz con tres elementos, dos de ellos designados.

Ésa es una razón para no estar satisfechos con $\mathbf{A}_\xi$. El fondo del problema

estriba en que no podemos con el limitadísimo vocabulario de $\mathbf{A}_\xi$ expresar ninguna relación más estrecha entre dos enunciados consistente en que uno de ellos sea más verdadero que el otro. Vamos ahora a partir de una ξ-matriz y vamos a añadir una operación binaria, I, como sigue. En primer lugar sólo tomamos ξ-matrices que tengan un elemento ½ tal que ½=N½. La operación I será tal que: xIz = ½ sys x=z; en caso contrario, xIz = 0. A una matriz así la llamaremos una ξI-matriz. (Podríamos generalizar ese tratamiento, exigiendo, en vez de igualdad, una congruencia *plenamente invariante*, o sea una que venga preservada por todos los endomorfismos.) El rasgo importante que añaden las ξI-álgebras y matrices es poder expresar la mismidad de grado de verdad. La lógica semánticamente definida como aquella cuyos modelos son ξI-matrices será la *lógica* $\mathbf{A}_\xi\mathbf{I}$. A diferencia de la lógica $\mathbf{A}_\xi$, $\mathbf{A}_\xi\mathbf{I}$ no es ni trivalente ni siquiera finivalente. Con respecto a una ξI-álgebra de n valores (n finito) se tendrá un esquema teoremático que no lo sea en $\mathbf{A}_\xi\mathbf{I}$; p.ej. en una lógica definida respecto a la clase unitaria de ξI-matrices cuyo portador es {0,½,1} se tendrá (definiendo ⌜p→q⌝ como ⌜p∧qIp⌝): ⌜p→(q∧Nq)∨¬q∨¬Nq∨¬Np⌝; y en general, para n valores se tendrá que, dadas n letras sentenciales, ⌜p^1⌝, ..., ⌜p^n⌝, y definiendo ⌜p\q⌝ como ⌜p→q∧¬(q→p)⌝, será teoremática la fórmula: ⌜p^1\\$p^2$∧($p^2$\\$p^3$)∧...($p^{n-1}$\\$p^n$)⊃.¬N$p^n$∧¬$p^1$⌝. (Dicho en plata: para cualquier cadena de n enunciados cada uno de los cuales sea más verdadero que los que lo precedan, el primero será totalmente falso y el último totalmente verdadero; ello excluye la utilización de lógicas finivalentes para el tratamiento lógico de los comparativos, según vino propuesto en Peña, 1987.) Igualmente, cada lógica finivalente contendrá como esquema teoremático una de las llamadas 'fórmulas de Dugundji': para n variables sentenciales, p^1, p^2, ..., p^n, la fórmula de Dugundji correspondiente es: ⌜p^1Ip^2∨.p^1Ip^3∨....∨.p^1Ip^n∨....∨.p^{n-1}Ip^n⌝. La fórmula de Dugundji en n variables sentenciales dice que hay un máximo de n-1 valores veritativos o grados de verdad. (La **LC** viene caracterizada por la fórmula de Dugundji en 3 variables: ⌜pIq∨.pIr∨.qIr⌝.) Ningún esquema así es teoremático en la lógica infinivalente $\mathbf{A}_\xi\mathbf{I}$.

La razón por la cual hemos tomado uniformemente xIz = ½ cuando x=z es la de poder así tener como teoremáticos *todos* los esquemas siguientes: ⌜pIq→.p∧rI.q∧r⌝; ⌜p→q∧p→q⌝; ⌜pIq→.rIqI.rIp⌝; ⌜pIq→.p∨rI.q∨r⌝; ⌜pIp→.qIq⌝; ⌜p→(p→q)→.p→q⌝; ⌜p→Np→Np⌝; ⌜Np→p→p⌝; ⌜p→q→.Nq→Np⌝; ⌜p→q→N(p∧Nq)⌝; ⌜p→q→.p→Nq→Np⌝; ⌜p→(q∧Nq)→Np⌝; ⌜p→q→.p→r→.p→.q∧r⌝; ⌜p→(q→r)→.p→q→.p→r⌝; ⌜p→q→.r→p→.r→q⌝; ⌜p→q→.q→r→.p→r⌝; ⌜N(p→p)→.p→p⌝; ⌜N(p→p→N(p→p))→.p→p⌝; ⌜p→q→r→(p→q)→.p→q⌝; ⌜p→q→r→.q→p→r→r⌝; ⌜p→q→(q→p)→.q→p⌝;

⌜p→q ∨ .q→p⌝ ; evitando, en cambio, la teorematicidad de los esquemas: ⌜p→.q→p⌝ ; ⌜p ∧ Nq→r→.p ∧ Nr→q⌝ ; ⌜p→(q→r)→.q→.p→r⌝ ; ⌜p→. p→q→q⌝ ; ⌜p→(q→.p→r)→.q→r⌝ ; ⌜p→.p→p⌝ ; ⌜p→q→.p→.p→q⌝ ; ⌜p→q→. q→.p→q⌝ ; ⌜p ∧ q→r→.p→.q→r⌝ ; ninguno de los cuales es compatible con la idea de que ⌜pIq⌝ sea una fórmula verdadera sys es tan verdadero que p como lo sea que q. (Es más: cada uno de esos esquemas no teoremáticos es tal que, escribiendo, para hacer las veces de su respectiva prótasis implicacional, ⌜p⌝ , y, para hacer las veces de su apódosis, ⌜q⌝ , no se cumple la condición: p ⊨q.) Los functores 'I' y '→' son interdefinibles: si '→' es primitivo, ⌜pIq⌝ es definible como ⌜p→q ∧ .q→p⌝ . De paso pruébanse otros esquemas característicos de otros sistemas lógicos, como el conexivismo: el llamado principio de Boecio (⌜p→q→N(p→Nq)⌝) y el de Aristóteles (⌜N(p→Np)⌝). Y finalmente estos dos: ⌜p→NpI.p→N(pIp)⌝ y el principio de Heráclito: ⌜N(pIp)⌝ .

La lógica $\mathbf{A}_{\xi}\mathbf{I}$ no es sólo paraconsistente sino contradictorial, ya que hay un cierto esquema tal que contiene como teoremas ese esquema y su negación simple: ⌜pIp⌝ y ⌜N(pIp)⌝ . También tenemos: ⌜pIpIN(pIp)⌝ .

Hay una razón importante para no estar todavía satisfechos con el resultado, y es que, si bien la lógica que hemos obtenido es genuinamente infinivalente (no es caracterizable por ninguna clase unitaria de matrices con un número finito de elementos), su vocabulario lógico es tan pobre que no podemos en ella expresar más que tres matices veritativos: o decir que ⌜¬p⌝ , o que ⌜¬Np⌝ o que ⌜p ∧ Np⌝ . De ⌜p ∧ Np⌝ y ⌜¬Nq⌝ podemos deducir ⌜p\q⌝ , pero nunca ⌜p\Np⌝ ni ⌜Np\p⌝ ni ⌜pINp⌝ . En general esta lógica no nos sirve como lógica de lo difuso porque no podemos establecer ninguna relación inferencial entre los matices de los asertos y la mayor o menor verdad de unos u otros. La lógica en cuestión no contiene ningún vocablo que exprese algo así como 'más bien', 'bastante', 'un tanto', 'muy', etc., ni, por lo tanto, teorema alguno que diga que, en la medida [al menos] en que algo sea muy verdadero, es verdadero [a secas]. (Eso de ⌜en la medida [al menos] en que p, q⌝ será nuestra lectura de ⌜p→q⌝ .)

No es ése el único motivo, como vamos a ver, para dar un paso más, introduciendo un nuevo functor primitivo. Otra razón es que hay ocurrencias de la conyunción 'y', o quizá más bien de otras conyunciones copulativas que no son semánticamente reducibles a ella, que no vienen adecuadamente capturadas por ' ∧ '. P.ej. hay un 'y' de insistencia —quizá mejor representado por la partícula discontinua 'no sólo ... sino [que] también'— en la cual parece que los conyuntos interactúan en el sentido de que el grado de falsedad resultante podrá ser mayor que los grados de falsedad de sendos conyuntos. Así, supongamos que una cierta oración, ⌜p⌝ , es verdadera en un

33% aproximadamente, mientras que ⌜r⌝ lo es en un 66%: según el tratamiento hasta aquí propuesto ⌜p y r⌝ será tan verdadera como ⌜p⌝, ni más ni menos; y lo propio sucederá para cualquier conyunción copulativa en vez de 'y'. Sin embargo cabe sospechar que al decirse ⌜p *y* r⌝ (o ⌜no sólo p, sino que además r⌝), se está diciendo algo menos verdadero que al decirse simplemente ⌜p⌝: porque ⌜r⌝ dista de ser del todo verdad, el aserto copulativo en cuestión ha de añadir algo más de falsedad al grado de falsedad que ya tenía ⌜p⌝. Representemos esa conyunción copulativa más fuerte como '●': aseverando ⌜p^1●p^2●p^3●...●p^n⌝, donde para cada $i \leq n$ ⌜p^i⌝ tiene un valor de verdad infinitamente inferior al máximo, se estará haciendo un aserto cuyo grado de falsedad será, *cæteris paribus*, tanto mayor cuantos más conyuntos haya (y no sólo cuanto menos verdaderos sean). La introducción de esa *superconyunción* nos va a permitir obtener, como definidos, muchos functores de matiz alético.

Otra razón más por la cual es conveniente añadir una conectiva que nos permita definir infinitos functores monádicos de matiz alético es ésta. El functor condicional '⊃' cumple los dos requisitos enumerados más arriba (rº1º) y (rº2º) para los condicionales, de suerte que podemos justificar la presencia de ese functor definido por su conexión con la deducción. En cambio nada similar justifica la presencia del functor implicativo '→': entendiendo la operación de consecuencia del modo clásico —que es el que ha venido adoptado en este trabajo (recuérdese la observación parentética del final del párrafo sexto de este mismo §2)—, no hay ningún nexo de inferencia entre $\{p^1,\ldots,p^n\}$ y $\{r\}$ suficiente para que sea teoremática la fórmula ⌜$p^1 \wedge \ldots \wedge p^n \rightarrow r$⌝. Dicho con otras palabras: el menor superconjunto de $\{p^1,\ldots,p^n\}$ cerrado con respecto a todas las reglas de inferencia de las lógicas que estamos examinando puede abarcar a ⌜r⌝ sin que por ello abarque a ⌜$p^1 \wedge \ldots \wedge p^n \rightarrow r$⌝. (El fundamento de ese desempate entre la inferibilidad y el functor implicativo '→' estriba en que el sentido de ⌜p→q⌝ es que ⌜q⌝ sea a lo sumo tan falso como lo sea ⌜p⌝, al paso que ⌜q⌝ se infiere de $\{p^1,\ldots,p^n\}$ sys o bien uno [al menos] de entre ⌜p^1⌝,..., ⌜p^n⌝ es del todo falso, o bien ⌜q⌝ es [en uno u otro grado] verdadero.) Esa falla puede corregirse con ayuda de la superconyunción y de los functores de matiz veritativo que mediante ella nos será dado introducir.

En el intervalo [0,1] podemos tomar x●z como el producto multiplicativo $x \times z$. Esta operación tiene los rasgos siguientes: conmutatividad, asociatividad, elemento neutro (el elemento máximo del álgebra en cuestión); además, ● es distributiva con respecto a las operaciones ∨ y ∧: x●(z∧u) = x●z∧(x●u); x●(z∨u) = x●z∨(x●u). (Eso significa que ● es una operación isótona, o sea que, si $x \leq z$, entonces x●u ≤ z●u.) Con respecto a N,

● tiene una característica especialmente importante, definiendo en general ⌜Ky⌝ como ⌜N(Ny●Ny)⌝ : si x = z●z, Kx = z; más en general: si x●x = z, u●u = v, entonces K(z●v) = x●u. Otra característica de ● es el *principio de cancelación*: x●z < x●u sys u>z.

Aunque hemos tomado como ejemplo un caso muy particular (esa matriz cuyo portador sea [0,1] y cuyo cúmulo de valores designados sea el filtro de elementos densos]0,1]), cabe señalar que hay muchas que son isomórficas con ésa. Pueden tomarse como ejemplos: primero el álgebra que llamaremos $\mathbf{A}_\infty$, a saber: una cuyo portador sea $\mathbb{R}$ = [—∞,+∞] donde ∧ sea la operación *max*, ∨ sea *min*, N sea —, ¬x=∞ si x≠∞, y ¬∞=—∞; y ● será así: ∞●x=∞=x●∞; —∞●x=x=x●—∞; si -∞<x≤z<∞, x●z = z●x = $\log_2(2^x+2^z)$ (+ aquí sí es la adición). Otra igual es un álgebra cuyo portador sea [0,∞], Nx sea 1/x, x●z sea x+z (+ también aquí es la adición), ∨ sea *min* y ∧ sea *max*: en estas álgebras ∞ es el elemento nulo o cero algebraico, y el orden algebraico es inverso al orden numérico usual. En la última álgebra considerada el ½ algebraico es el 1 numérico: en la anterior, el 0 numérico.

Como todas esas álgebras son isomórficas entre sí, tienen una serie de rasgos además de los que nos interesan. Veámoslo con un simple ejemplo: tomemos el intervalo de los números racionales [0,∞] con las operaciones definidas igual que sobre el intervalo de los reales con los mismos extremos. Se ve en seguida que no son isomórficas ambas álgebras.

Mientras que hay muchas lógicas que tienen matrices características finitas, hay pocas que tengan matrices características infinitas. Generalmente, cuando se transciende la finitud hay que caracterizar a un cálculo lógico por una clase entera de matrices —con múltiples, en verdad infinitos, miembros— y ya no por una matriz en particular (salvo las matrices de Lindenbaum, que son matrices cuyos respectivos portadores son clases de fórmulas).

Hay un importantísimo rasgo de la clase de aquellas ξI-matrices a las que se haya enriquecido con la operación ●: ninguna de tales matrices tiene un número finito de elementos; y ello por el postulado de cancelación.

Hay todavía una razón para pensar que está incompleta nuestra busca de operaciones: llamemos matrices Ω a las que están algebraicamente caracterizadas por las operaciones (∨, ∧, ●, I, N y ¬) y por los postulados que acabamos de proponer, siendo en cada caso *D* el filtro de los elementos densos, o sea de los elementos x tales que ¬x = 0. Salta a la vista que en muchas de estas matrices 0 es el ínfimo de *D* (el elemento *ínfimo* de un subconjunto X de un conjunto Z —notado como ⋀X— es la mayor cota inferior

de X, donde x es cota inferior de X sys cada $z \in X$ es tal que $z \geq x$; dualmente, el elemento *supremo* de X, $\bigvee X$, es su menor cota superior, donde x es cota superior de X sys no hay $z \in X$ tal que $z \nleq x$). Eso es muy grave, porque para extender un cálculo sentencial a uno cuantificacional habrá que dar al cuantificador un tratamiento más o menos así: una valuación, v, enviará a la fórmula $\ulcorner\forall xp\urcorner$ sobre una imagen suya, u, sys u es el ínfimo del conjunto de elementos z tales que $z = v'(\ulcorner p\urcorner)$ para alguna x-variante v' de v, donde una valuación es una x-variante de otra sys a cualquier argumento que no contenga la variable 'x' le hacen corresponder la misma imagen (el mismo valor). Entonces puede suceder que para cierta fórmula $\ulcorner p\urcorner$ se tenga que sea válida cada fórmula $\ulcorner p[x/a]\urcorner$ (cada resultado de reemplazar uniformemente en $\ulcorner p\urcorner$ las ocurrencias libres de 'x' por sendas ocurrencias de un término 'a'), pero en cambio $\ulcorner\forall xp\urcorner$ sería inválida. Basta para ello con tomar como $\ulcorner p\urcorner$ el principio fuerte de tercio excluso, $\ulcorner r \vee \neg r\urcorner$, siendo $\ulcorner r\urcorner$ una fórmula con alguna ocurrencia libre de la variable 'x', pero por lo demás indeterminada por los postulados que hemos sentado. Es obvio que para cada valuación v que envíe a $\ulcorner r\urcorner$ sobre un elemento denso, u, habrá otra que lo envíe sobre otro $z < u$. Entonces, si bien será una fórmula válida $\ulcorner r \vee \neg r\urcorner$, no lo será $\ulcorner\forall x(r \vee \neg r)\urcorner$. (Es más: ¡$\ulcorner\neg\forall x(r \vee \neg r)\urcorner$ será válida!) La regla de generalización universal ($p \vdash \forall xp$) no será preservadora de la validez, ni siquiera del estatuto de designación (no se tendría, pues, $p \models \forall xp$).

Un remedio contra ese mal sería redefinir nuestras matrices de suerte que el filtro de los elementos designados no abarcara a todos los elementos densos, siendo el ínfimo de dicho filtro miembro del filtro. Pero entonces perderíamos el principio fuerte de tercio excluso y la definibilidad clásica del condicional, '$\supset$', $\ulcorner p \supset s\urcorner$ como abreviación de $\ulcorner\neg p \vee s\urcorner$. (Eso se debería a que habríamos perdido también la regla del silogismo disyuntivo para la negación fuerte, a saber: $p \vee \neg q,\ q \models p$.) También perderíamos el principio fuerte de no-contradicción, a saber: $\ulcorner N(Lp \wedge Np)\urcorner$, donde $\ulcorner Lp\urcorner$ abrevia a $\ulcorner N\neg p\urcorner$. Al tener que abandonar el condicional '$\supset$' definible del modo indicado, perderíamos muchas propiedades del condicional clásico, que es lo que les pasa a todas las lógicas multivalentes en las que se toma como filtro de los elementos designados a un subconjunto propio del cúmulo de los ele-

mentos densos de un álgebra de Stone. Perderíamos asimismo lo que cabe llamar *regla de apencamiento*, a saber: Lp ⊨p. En suma, caeríamos en una lógica más pobre, en vez de tener una que sea más rica que **LC**. (Lo que sí es cierto es que las primeras construcciones, históricamente, de lógicas multivalentes, reseñadas más arriba —en el §1— siguieron ese camino; ninguna de aquellas lógicas era, pues, una extensión conservativa de **LC**.) Por otra parte, aun estando dispuestos a irnos al extremo de la parsimonia o austeridad y a no tomar como elemento designado más que al 1 algebraico o elemento máximo (que es lo que efectivamente hicieron las primeras lógicas multivalentes), no se habrían acabado las dificultades de esta índole: de hecho —según vino ya indicado en el §1— el cálculo cuantificacional infinivalente de Łukasiewicz es inaxiomatizable.

En lugar de seguir ese camino trillado de achicamiento del conjunto de elementos designados, vamos a explorar otro, consistente en transformar a las álgebras que nos interesan en *retículos fuertemente algebraicos*, definidos como sigue. Un retículo es *completo* sys en él cada subconjunto de su portador (¡aun uno vacío!) tiene un ínfimo y un supremo. Obviamente para poder extender el cálculo sentencial a un cálculo cuantificacional hace falta que los modelos algebraicos que consideremos sean retículos completos. Pero no basta, ya lo hemos visto. (Si tomamos un álgebra cuyo portador sea un intervalo de números racionales, sería incompleta, pero si es de números reales será completa.) Hace falta que el filtro de los valores de verdad designados sea un *filtro completo*, o sea uno F tal que, si G ⊆ F, ⋀G ∈ F. Vamos a ver que una condición que asegura eso es que se trate de un retículo fuertemente algebraico.

Dícese que en un retículo z cubre a x, x⊸z o z⟜x, sys z > x y no hay ningún elemento u tal que z > u > x. Un retículo será llamado *atómico* sys para cada par de intervalos contiguos [a,b] y]b,c] (o sea cada par de subconjuntos {v: a≤v≤b} y {v: b<v≤c}) hay un elemento v ≠ b tal que v⊸b o b⊸v; o sea: o v es el ínfimo del segundo intervalo, o es el supremo del intervalo [a,b[. Un elemento x es *compacto* sys todo subconjunto S de su portador es tal que: si ⋀S ≤ x, entonces x ≥ ⋀S', siendo S' un subconjunto finito de S. Un retículo es *fuertemente algebraico* (o, con otra palabra, *engen-*

drado de modo fuertemente compacto) sys es completo y cada par de intervalos cuasicontiguos [a,b[y]b,c] es tal que o bien el ínfimo del segundo es compacto, o bien lo es el supremo del primero.

Aunque esas nociones pueden parecer abstrusas, su utilidad estriba en este importantísimo **Teorema**: todo retículo fuertemente algebraico es atómico. Prueba: sean en el retículo fuertemente algebraico L tres elementos, $a < e < b$, sin que se tenga ni $a \lessdot e$ ni $e \lessdot b$ (de tenerse eso, tendríamos de entrada lo que buscamos). Tomemos los dos intervalos cuasicontiguos [a,e[y]e,b]. Supongamos que es compacto el supremo del primer intervalo, d. Tenemos que $d \leq \bigwedge]d,b]$. Supongamos que no sólo es $\leq$ sino que es también $\geq$, o sea que $d = \bigwedge]d,b]$, lo cual implica que $d=e$. Por ser compacto, habrá un subconjunto finito G de]d,b] tal que $d \geq \bigwedge G$, cosa imposible (G es finito, luego su ínfimo ha de ser miembro de G). Por el lema de Zorn se concluye que]d,b] tiene un elemento minimal. (El lema de Zorn reza: si cada cadena de un conjunto [parcialmente] ordenado no vacío, E, tiene una cota inferior, entonces E posee un elemento x minimal, e.d. uno $z \in E$ tal que para todo $x \in E$ $x \not< z$.) Si ese elemento minimal es e, la hipótesis era falsa: $d \neq \bigwedge]d,b]$, o sea $d \lessdot e$, pues evidentemente no puede haber ningún elemento entre d y e. Mas, si ese elemento minimal no es e, entonces tenemos lo que buscábamos: $e \lessdot \bigwedge]eb]$. Supongamos entonces la otra alternativa, o sea que $\bigwedge]e,b] = h$ es compacto. Pruébase exactamente igual entonces que $h < \bigwedge]h,b]$, o sea que existe un elemento k que es $\bigwedge]h,b]$ y tal que $h \lessdot k$. Si $h=e$, ya tenemos lo que andábamos buscando. Mas si no también, porque si $h \neq e$, $e \lessdot h$. Q.E.D.

Otro teorema dice que todo retículo fuertemente algebraico es algebraico, siendo un *retículo algebraico* uno en el que cada miembro es el ínfimo de un determinado conjunto de elementos compactos. He aquí la prueba: sea en un retículo fuertemente algebraico R un elemento e que no cumpla la condición. e estará en una línea divisoria entre intervalos cuasicontiguos [a,e[y]e,b], para ciertos a y b. Tomemos el intervalo]e,b]; cada miembro de ese intervalo tendrá «al lado» (por arriba o por abajo) un elemento compacto; sea G el conjunto de esos infinitos elementos compactos pertenecientes al intervalo]e,b]. La hipótesis que tratamos de reducir al absurdo nos fuerza a afirmar

que $\bigwedge G > e$. Mas eso no es posible, porque entonces habrá sólo un número finito de elementos entre e y $\bigwedge G$, y entonces el propio e sería compacto (y por ende sería $\bigwedge\{e\}$). Q.E.D.

Suele llamarse *noetherio* a un retículo que no contenga ninguna serie infinita de elementos, $a^1, \ldots, a^n, a^{n+1}, \ldots$, tales que $\ldots > a^n > a^{n-1} > \ldots > a^2 > a^1$. (Por el axioma de elección se prueba que tal condición equivale a que todo subconjunto no vacío del portador del retículo tenga un elemento minimal.) Pruébase con facilidad que todo retículo noetherio es algebraico. De ahí que se hayan buscado como modelos para la mayor parte de las lógicas multivalentes retículos noetherios, y por lo tanto que sean o finitos o productos de retículos finitos. Y es que, para su aplicación a los cuantificadores, la característica de que sea algebraico el retículo de los valores de verdad parece más que un mero *desideratum*. Además de la razón ya considerada (y decisiva), referida a la regla de generalización universal, existe otra, y es que hay un teorema del álgebra universal que reza así: todo retículo algebraico es *continuo*, entendiendo por tal uno con esta condición: para cada elemento b y cada cadena C, $b \vee \bigwedge C = \bigwedge\{b \vee z: z \in C\}$. Para el tratamiento de los cuantificadores esa condición de continuidad significa la aplicabilidad de lo que se llama *ley de paso*: si ⌜p⌝ no contiene ninguna ocurrencia libre de la variable 'x', ⌜$p \vee \forall x q I \forall x(p \vee q)$⌝ es teoremático; esa condición y la dual respectiva permiten aplicar procedimientos de prenexación, desprenexación, conversión a forma normal, etc.

Ya hemos visto empero los inconvenientes de la finitud. Es arbitrario postular que se tengan que dar exactamente n grados de verdad en vez de n+1. No hay números finitos lógicamente privilegiados. Afortunadamente hay cómo obtener que un retículo sea fuertemente algebraico (y por lo tanto atómico y continuo, en sendos sentidos más arriba apuntados) sin incurrir en la finitud. Es lo que voy a exponer a continuación.

Partimos del álgebra $\mathbf{A}_\infty$, según vino definida más arriba (usando aquí los signos '$\leq$' y '$<$' para referirnos al orden numérico, inverso del algebraico), pero ensanchando su portador a un conjunto $S = \mathbb{R} \cup \{<r,a>: r \in \mathbb{R} \ \& \ r > -\infty\} \cup \{<r,b>: r \in \mathbb{R} \ \& \ r < \infty\}$, donde a, b son entes cualesquiera. Postulamos: $<r,a> \ < \ r \ < \ <r,b>$. Y definimos (dejando correr a las variables 'r', 'r^1', 'r^2' sobre miembros de $\mathbb{R}$ [incluidos ∞ y $-\infty$]):
$<r^1,a> \bullet <r^2,b> \ = \ <r^1 \bullet r^2,b>$ si $r^1 \neq \infty$; $<\infty,a> \bullet x \ = \ <\infty,a>$ si $x \neq \infty$;

$<r^1,a>\bullet r^2 = <r^1,a>\bullet<r^2,a> = <r^1\bullet r^2,a>$ si $r^2\neq\infty$; $\infty\bullet x=\infty$; $<r^1,b>\bullet r^2 = <r^1,b>\bullet<r^2,b> = <r^1\bullet r^2,b>$ si $r^2\neq\infty$; $-\infty\bullet x=x$; y siempre $x\bullet z = z\bullet x$ (lo cual termina de definir recursivamente la operación). Las definiciones de las operaciones $\vee$, $\wedge$, $\neg$ son las mismas que en $\mathbf{A}_\infty$; la operación N viene extendida así: $N<r,a> = <Nr,b>$ y $N<r,b> = <Nr,a>$. A esta álgebra la llamaremos $\mathbf{A}_{\omega\alpha}$; al elemento $<-\infty,b>$ lo llamaremos ω; a $<\infty,a>$ llamarémoslo α.

En esta álgebra, el filtro de los elementos designados es $\{x: x\geq\alpha\}$, donde '$\geq$' es el orden algebraico (inverso al numérico). Podemos también introducir el *ideal* de los elementos antidesignados (un ideal es un conjunto C tal que, si $x\leq z$, $z\in C$ entraña $x\in C$, y $x\vee z\notin C$ sólo si o bien $x\notin C$ o bien $z\notin C$); será tal ideal $\{x: x\leq\omega\}$. Todos esos signos encuentran lecturas naturales en la lengua vernácula: '$\neg$' : 'No...en absoluto'; 'N': 'no'; '$\vee$': 'o'; '$\wedge$': 'y'; '$\bullet$': 'no sólo... sino que también'; '$\supset$': 'sólo si'; '$\to$': 'sólo en la medida en que' o 'sólo en tanto en cuanto'; 'I': 'en la misma medida en que'; $-\infty$ es lo totalmente verdadero; ∞ la falsedad completa; α la verdad infinitesimal; ω el grado infinitesimal de falsedad (que es un grado infinito, aunque no total, de verdad).

Tomemos un lenguaje cuyas valuaciones tomarán sus valores en $\mathbf{A}_{\omega\alpha}$. Definiendo $\ulcorner np\urcorner$ como $\ulcorner p\bullet\omega\urcorner$ y $\ulcorner mp\urcorner$ como $\ulcorner NnNp\urcorner$, la primera fórmula cabe leer como «Es supercierto que p», o algo así, y la segunda como «viene a ser cierto que p». Definiendo $\ulcorner Pp\urcorner$ como $\ulcorner\neg\neg(Np\to p)\wedge p\urcorner$ (que cabe leer como «Es más bien cierto que p»: 'P' por 'potius'), serán válidos los esquemas $\ulcorner Pp\vee PNp\urcorner$, $\ulcorner p\to q\to .Pp\to Pq\urcorner$. Valdrán las inferencias $p\dashv\;\vDash\alpha\to p$, $Np\dashv\;\vDash p\to\omega$, $\neg p\dashv\;\vDash p\to .q\wedge\neg q$ y $\neg p\dashv\;\vDash p\backslash\alpha$. Definiendo $\ulcorner Yp\urcorner$ como $\ulcorner pI\alpha\wedge p\urcorner$, cabe leerlo como «Es infinitesimalmente verdad que p», y se tiene: $Yp\dashv\;\vDash p\wedge\neg(\alpha\backslash p)$.

En $\mathbf{A}_{\omega\alpha}$ no vale ya el principio de cancelación que valía en $\mathbf{A}_\infty$, salvo con una modificación, a saber: si $x\bullet z\approx u\bullet z$, entonces o bien $z\approx\infty$ (el cero algebraico), o bien $x\approx u$, donde '$\approx$' expresa una diferencia a lo sumo infinitesimal de grado (o sea: $<r,b> \approx r \approx <r,a>$, pero en ningún otro caso $x \approx z$). $\approx$ no es una congruencia. Pero cabe generalizar la noción de congruencia así: para el conjunto de operaciones $\{♪_i\}_{(i\in I)}$ se tendrá que una relación de equivalencia, Θ, es una $\{♪_i\}_{(i\in I)}$-congruencia sys tiene la propiedad de sustitución para cada $♪_i$ $(i\in I)$. Pues bien, $\approx$ es una $\{\wedge,\vee,N,\bullet,m\}$-congruencia. (No tiene la propiedad de sustitución ni para I ni para $\neg$.)

En la semántica aquí esbozada $p\dashv\;\vDash q$ no será una congruencia en el cálculo semánticamente definido (de $p\dashv\;\vDash q$ no se sigue $Np\dashv\;\vDash Nq$, ni

$Yp\dashv\models Yq$ ni $Pp\dashv\models Pq$, etc.). Mas sí será una $\{\wedge, \vee, \neg, \bullet\}$-congruencia. En verdad se trata de la relación de equivalencia entre $\ulcorner p\urcorner$ y $\ulcorner q\urcorner$ sys $\models \neg\neg p\mathrm{I}\neg\neg q$, la cual «corresponde» a su vez a la congruencia de Glivenko para un álgebra de Stone, a saber: $x\Theta_G z$ sys $\neg\neg x=\neg\neg z$. El fragmento de este cálculo que tiene como constantes lógicas '$\wedge$', '$\vee$' y '$\neg$' es exactamente **LC**. (El cálculo, como se indicó en el §1, es además una extensión cuasiconservativa de cada lógica finivalente.)

Podemos añadir a las operaciones de esta álgebra un conjunto infinito de operaciones unarias δ_s, $s\in S$—$\{\infty\}$, donde se tendrá: $\delta_s x=x$ si $s=x$; y, si no, $\delta_s x=\infty$. Entonces tendremos en el cálculo semánticamente definido con relación a esta álgebra sendos functores δ_s; $\ulcorner\delta_s p\urcorner$ dice que es verdad que p en grado s. Resultado: sys $\delta_s p\dashv\models\delta_s q$ para todo s, $\models p\mathrm{I}q$. Y, más en general, sys para todo s $X\models\delta_s p$ sys $X\models\delta_s q$, $X \models p\mathrm{I}q$. Trátase, naturalmente, de secuentes infinitarios.

Ya hablamos más arriba de la relación entre las lógicas multivalentes y las relevantes. Precisamente una lógica que admite una bonita definición semántica como lógica infinivalente es el sistema (de la familia relevante) **RM**. (Cuando se trata de sistemas relevantes, es a menudo difícil saber quién es el originador de uno de ellos en particular; mas lo que es seguro es que quien más se ha destacado en el estudio de **RM** es Robert K. Meyer; v. su sección sobre ese sistema, §29.3, de Anderson y Belnap, 1975, 393 ss.) He aquí la matriz característica de dicho sistema: $<\mathbf{I}, \mathbf{D}, N, \wedge, \vee, \rightarrow>$, donde $\mathbf{I}$ es el conjunto de los enteros; $\mathbf{D}$ es el conjunto de los enteros no positivos; $i\wedge j = max(i,j)$; $i\vee j = min(i,j)$; $Ni = -i$; y $j\rightarrow i =: Nj\vee i$ si $j\geq i$, y $Nj\wedge i$ si $j<i$. Este sistema tiene una axiomatización finita muy elegante que resulta de añadir al sistema «relevante» **R** el célebre axioma «Mingle»: $\ulcorner p\rightarrow.p\rightarrow p\urcorner$. Sin embargo, **RM**, aunque de esa familia, ya no es un sistema que cumpla los constreñimientos relevantes.

Pues bien, recientemente se ha estudiado una serie de sistemas intermedios entre las lógicas de la familia **A** —la que se ha venido exponiendo someramente en párrafos precedentes de este §— y el sistema relevante **E**, que resulta de **R** al suprimir el axioma de permutación: $\ulcorner p\rightarrow(q\rightarrow r)\rightarrow. q\rightarrow.p\rightarrow r\urcorner$. He aquí cómo es uno de ellos, **P5**. Puede axiomatizarse con estos 10 esquemas axiomáticos:

$p\rightarrow q\rightarrow r\wedge(q\rightarrow p\rightarrow r)\rightarrow r$

$p\rightarrow q\wedge(q\rightarrow r)\rightarrow.p\rightarrow r$

$p\wedge q\wedge r\rightarrow.r\wedge p\wedge q$

$p\rightarrow q\rightarrow.r\rightarrow s\rightarrow.p\rightarrow q\wedge.r\rightarrow s$

$p\rightarrow q\rightarrow r\rightarrow s\wedge(\blacklozenge(p\rightarrow p)\rightarrow(p\rightarrow q)\rightarrow s)\rightarrow s$

$p\wedge q\rightarrow p$

$p\rightarrow q\rightarrow.p\rightarrow.p\wedge q$

$Np\rightarrow q\rightarrow N(p\rightarrow q)$

p→Nq→.q→Np NNp→p

Los functores primitivos son '∧', '→', 'N'. '♦' es definido así: ⌜♦p⌝ abr. ⌜N(p→Np)⌝; ⌜p∨q⌝ abr. ⌜N(Np∧Nq)⌝. Una sola regla de inferencia: para $n \geq 1$: ⌜p^1→q∨.p^2→q∨....∨.p^n→q⌝, ⌜p^1⌝, ..., ⌜p^n⌝ ⊢ q. (Cuando $n=1$, trátase del *modus ponens* normal y corriente.) Este sistema tiene como característica una clase de matrices infinitas, parecidas a la de **RM** (y de hecho varias similitudes con **RM**): el portador será $I \cup \{-\infty, \infty\}$; las mismas asignaciones para los functores excepto para '→': j→i =: ∞ si $j<i$; 0 si $j \geq i$; el conjunto de elementos designados, **D**, variará de una matriz a otra, si bien siempre $[-\infty, 0] \subseteq$ **D**. Esa clase de matrices se llamará **P**, siendo P_j aquella matriz de la clase **P** cuyo conjunto de elementos designados sea $\{x: x \leq j\}$; y $P_{<j}$ aquella en que sean designados todos los elementos < j. Las discrepancias entre **RM** y **P5** son tan significativas como las convergencias. P.ej. en **RM**, mas no en **P5** vale este teorema: ⌜p∧Np∧q∧Nq→.pIq⌝ (donde, claro, ⌜pIq⌝ abr. ⌜p→q∧.q→p⌝). **P5** y **RM** son paraconsistentes, mas sólo el primero es contradictorial.

¿Tiene **P5** una matriz característica? No puede ser $P_{<\infty}$, porque en ella para cualesquiera fórmulas, ⌜p⌝, ⌜q⌝, se tendrá la validez (con relación a $P_{<\infty}$, tomada como característica) de la fórmula ⌜p→q∨p⌝, que no es teoremática en **P5**. Por otro lado, hay una diferencia entre **P5** y una lógica definida como caracterizada por la matriz P_0: en esa lógica la operación de consecuencia $\models_P$ será tal que {p, Np, q, Nq} $\models_P$ pIq (igual que en en **RM**). Y esa regla de inferencia no es derivable en **P5**. Sin embargo, los teoremas de una lógica así definida sí son exactamente los de **P5**.

El más fuerte de los sistemas de esa cadena es **P10**, un sistema que el lector va a reconocer inmediatamente como de la familia **A**, que hemos venido examinando; **P10** añade a **P5** dos nuevos functores, 'H' monádico ('Es totalmente cierto que') y 'α', 0-ádico, o sea una constante sentencial, que denota la conyunción de todas las verdades (lo infinitesimalmente verdadero). Los esquemas axiomáticos adicionales son cuatro:

p→q→.NHNHp→Hq Hp→q∨.Np→r α→p∨.p→q α

P10 sí tiene una matriz característica, que es como una para **P5** sólo que el portador ha sido aumentado con dos elementos, α (siendo cualquier entero e < α < ∞) y Nα, tal que cualquier entero e > Nα > −∞. Todos los elementos son designados salvo ∞. Cualquier valuación, *v* será tal que *v*(Hp) = −∞ si *v*(p) = −∞, y, si no, *v*(p) = ∞. *v*(α)=α (para cualquier *v*). Todo eso es como el lector ya lo esperaba, a estas alturas.

Quedan todavía por resolver muchísimos problemas con respecto al género de lógicas infinivalentes cuyo estudio viene posibilitado por las álgebras del tipo recién considerado. P.ej.: ¿cuál es la axiomática más simple y elegante para el álgebra $\mathbf{A}_{\omega\alpha}$ y las a ella isomórficas? Una vez dados ciertos postulados que recojan las propiedades lógicamente interesantes de esas álgebras (y se han propuesto varios conjuntos de tales postulados, recogidos en las obras ya citadas, que figuran en la bibliografía), ¿cuál es la menor álgebra —que no sea simplemente un álgebra de Lindenbaum o de Tarski— que los satisfaga todos (y que, por lo tanto, sea un *retracto* de $\mathbf{A}_{\omega\alpha}$, donde un álgebra A es un retracto de otra B sys hay un automorfismo idéntico de A [uno, m, tal que mx=x siempre] que es la composición o producto relativo de un monomorfismo de A en B con un epimorfismo de B en A)?

III CONCLUSIONES

Están aún por investigar las cuestiones con que ha finalizado el § precedente —así como muchas otras—, pero lo que ya parece probado es que ese género de tratamiento abre perspectivas que incrementan la aplicabilidad y el grado de motivación filosófica de las lógicas multivalentes. De hecho, ese manido aserto de que las lógicas multivalentes son meros juegos matemáticos ha sido siempre desacertado (ya Łukasiewicz puso en pie su sistema movido por una idea filosófica, equivocada o no, que es el rechazo del determinismo), pero nunca ha sido tan falso como con relación a las lógicas algebraicas infinivalentes que acabamos de esbozar.

La idea de que hay sólo dos valores de verdad es tan respetable como cualquier otra tesis metafísica, añeja o no, pero frente a ella abonan razones de peso que no cabe dejar de escuchar atentamente; algunas de esas razones llevaron a una parte de la tradición filosófica —aunque minoritaria— a la afirmación de grados de realidad y de verdad; otras de tales razones tienen que ver con problemas epistemológicos debatidos actualmente; y muchas de ellas guardan conexión con aplicaciones de la lógica a diversos campos del saber y de la investigación.

Teniendo en cuenta que generalmente el mundo se nos acaba presentando como más complicado de lo que nos lo solíamos imaginar, cabe conjeturar que es infinitamente complicado, y que una parte de esa complejidad viene dada por la infinivalencia veritativa, por los infinitos grados de verdad y de falsedad. También habría que tener en cuenta otra faceta, que multiplica al

infinito la infinitud misma: en este trabajo sólo hemos considerado lógicas escalares, salvo una breve alusión a las álgebras producto. Hay razones —en las que ya no cabe entrar aquí— para pensar que la realidad es más complicada, y que incurren en simplificación burda las lógicas escalares (aquellas en las que, para cualesquiera dos valores, x, z, $x \leq z$ o $z < x$): sería, en tal caso, más correcto representar a los valores de verdad como tensores o matrices infinitas (en el sentido del cálculo matricial). Entre otras cosas, así se podría dar un tratamiento más adecuado a problemas como algunos de la física cuántica o los del realismo modal de David Lewis y temas afines.

Una repercusión de la adopción de una lógica así sería que habría teorías aceptables no primas, o sea tales que $\ulcorner p \vee q \urcorner$ podría ser afirmable con verdad sin que lo fueran ni $\ulcorner p \urcorner$ ni $\ulcorner q \urcorner$. Naturalmente eso acarrea ciertas complicaciones para el tratamiento semántico de la disyunción.

Lo que parece más allá de la controversia es que las lógicas multivalentes no son ni matemática ni filosóficamente anodinas.

Bibliografía

Anderson, A.R. & Belnap Jr., N.D. (1975), *Entailment: The Logic of Relevance and Necessity*, Princeton University Press, Princeton.

Arruda, A.I. y ot. (comps.), *Proceedings of the Third Brazilian Conference on Mathematical Logic*, Universidade Estadual de Campinas, Campinas (São Paulo).

Balbes, R. & Dwinger, P. (1974), *Distributive Lattices*, University of Missouri Press, Columbia (Missouri).

Barba Escribá, J. (1992), «Lógica clásica, parcial y tetravaluada: una visión comparativa desde el punto de vista algebraico». Aparecerá en las Actas del Encuentro de lógica y filosofía de la ciencia, Madrid nov. 1991.

Belnap Jr., N.D. (1977), «A Useful Four-Valued Logic», en: Dunn & Epstein (comps.) (1977).

Cignoli, R. (1980), «Some Algebraic Aspects of Many-valued Logics», en: Arruda y ot. (comps.), 49-70.

Costa, N.C.A. da, Subrahmanian, V.S. & Vago, C. (1989), «The paraconsistent logics *PT*». Universidade de São Paulo, São Paulo (prepublicación).

Dunn, J.M. & Epstein, G. (comps.) (1977), *Modern Uses of Multiple-Valued Logic*, Reidel, Dordrecht.

Gabbay, D. & Guenthner, F., (comps.) (1986), *Handbook of Philosophical Logic*.

Guccione, S., & Tortora, R. (1982), «Deducibility in Many-Valued Logics», en: IEEE comps (1982), 117-121.

Haack, S. (1974), *Deviant Logic: Some Philosophical Issues*, Cambridge University Press, Cambridge.

IEEE (comps.) (1982), *The Twelfth International Symposium on Multiple-Valued Logic*, IEEE Computer Society Press, Silver Spring (Maryland). Son las Actas del XII congreso de lógica multivalente, celebrado en París, en mayo de 1982.

Łukasiewicz J. (1967), «On Three-Valued Logic», en: McCall (comp.) (1967), 16-18.

Malinowski, G. (1979), *Topics in the Theory of Strengthenings of Sentential Calculi*, Institute of Philosophy and Sociology, Varsovia.

Martín Vide, C. (comp.) (1992), *Lenguajes naturales y lenguajes formales* **VII**. PPU (Promociones y Publicaciones Universitarias), Barcelona.

McCall, S. (comp.) (1967), *Polish Logic: 1920-1939*, Clarendon, Oxford. (Con introducción de T. Kotarbiński y traducción de los originales polacos.)

Méndez, J.M. (199?), «Routley-Meyer Type Semantics for Urquhart's *C*», *Journal of Non-Classical Logic*, en vías de publicación.

Moisil, G. (1972), *Essais sur les logiques non chrysippiennes*, Éditions de l'Académie de la République Socialiste de Roumanie, Bucarest.

d'Ottaviano, Í.M. (1982), *Sobre uma teoria de modelos trivalente*, Universidade Estadual de Campinas, Campinas (São Paulo).

Peña, L. (1987), «Contribución a la lógica de los comparativos», en: *Lenguajes naturales y lenguajes formales* **II**. Universitat de Barcelona, Barcelona, 335-50. Compilado por Carlos Martín Vide.

Peña, L. (1991), *Rudimentos de lógica matemática*, CSIC, Madrid.

Peña, L. (1992) «Nuevos avances en la articulación y en las aplicaciones de lógicas aléticas», en Martín Vide (comp.) (1992), págªs 209-220.

Peña, L. (1994), *Introducción a las lógicas no clásicas*, UNAM, México.

Rasiowa, H. (1974), *An algebraic Approach to Non-classical Logics*, North-Holland, Amsterdam.

Rautenberg, W. (1979), *Klassische und nichtklassische Aussagenlogik*, Friedr. Vieweg & Sohn, Braunshwaig/Wiesbaden.

Rescher, N. (1969), *Many-valued Logic*, McGraw-Hill, Nueva York.

Rosser, J.B. & Turquette, A.R. (1952), *Many-Valued Logics*, North-Holland,

Amsterdam.

Sylvan, R. & Urbas, I. (1989), *Factorisation Logics*, Australian National University, Canberra. Monografía Nº 5 de la Research Series in Logic and Metaphysics.

Trillas, E. & Valverde, L (1982), «A Few Remarks on Some Lattice-Type Properties of Fuzzy Connectives», en: IEEE (comps.) (1982), 228-231.

Urquhart, A. (1986), «Many-Valued Logic», en: Gabbay & Guenthner comps (1986), vol. III, 71-116.

Varlet, J. (1975), *Structures algébriques ordonnées*, Université de Liège, Institut de Mathématique, Lieja.

Wright, H. von (1987), «Truth Logics», *Logique et Analyse*, **30**, 311-334.

Zadeh, L. y ot. (comps.) (1975), *Fuzzy Sets and their Application to Cognitive and Decision Processes*, Academic Press, Nueva York.

INDICE ANALÍTICO

INDICE DE NOMBRES

NOTA BIOGRÁFICA DE AUTORES

Carlos Eduardo Alchourrón (Buenos Aires, Argentina, 1931), es especialista en lógica y en filosofía del derecho, desarrollando su labor académica en la Universidad de Buenos Aires. Autor, en colaboración con E. Bulygin, de *Normative Systems* (1971) y *Análisis lógico y Derecho* (1991).

Eugenio Bulygin (Colomb, Rusia, 1931) ejerce la docencia en la Universidad de Buenos Aires, donde se ha especializado en filosofía del derecho. Es autor, entre otras, de las siguientes obras: *Naturaleza jurídica de la letra de cambio. Un ensayo de análisis lógico de conceptos jurídicos* (1961), *Normative Systems* (en colaboración con C. Alchourrón) (1971), *Introducción a la metodología de las ciencias jurídicas y sociales* (1975), *Sobre la existencia de las normas jurídicas* (1979), *Análisis lógico y Derecho* (1991).

Raúl Jorge Carnota (Buenos Aires, Argentina, 1949), especialista en inteligencia artificial, lógicas no-monótonas y revisión de creencias, ejerce la docencia en la Facultad de Ciencias Exactas de la Universidad de Buenos Aires. Es autor, entre otras, de las siguientes publicaciones: *Inteligencia Artificial aplicada* (1986), *Sistemas expertos y representación del conocimiento* (1988), *Non monotonic Preferential Models and Conditional Logic* (1989), *Modus Ponens condicional y lógicas no-monótonas* (1992).

Newton C. A. da Costa (Curitiba, Brasil, 1929), profesor del Departamento de Filosofía en la Universidad de São Paulo donde imparte enseñanzas de lógica, teoría de la ciencia y fundamentos de la física. Autor de numerosos artículos en revistas y medios especializados así como de las siguientes obras: *Mathematical Logic* (1978), *Mathematical Logic in Latin America* (1980), *Lógica Indutiva e Probabilidade* (1993), *Essai sur les Fondements de la Logique* (1994).

Max Freund Carvajal (Costa Rica, 1954), es profesor del Departamento de Filosofía en la Universidad Nacional de Costa Rica y experto en lógica modal y lógica de segundo orden. Entre sus publicaciones en medios especializados se encuentran: *Consideraciones lógico-epistémicas relativas a una forma de conceptualismo ramificado* (1991), *Un sistema lógico de segundo orden conceptualista con operadores lambda ramificados* (1992), *The relative consistency of System RRC* and some of its extensions* (1994).

Ignacio Jané Palau (Tarragona, España, 1945), es profesor del Departamento de Lógica, Historia y Filosofía de la Ciencia en la Universidad de Barcelona. Entre sus publicaciones se hallan: *Álgebras de Boole y Lógica* (1989), *A critical appraisal of second-order logic* (1993).

Sergio F. Martínez Muñoz (Ciudad de Guatemala, Guatemala, 1950), es investigador de historia y filosofía de las ciencias naturales en el Instituto de Investigaciones Filosóficas de la Universidad Nacional Autónoma de México. Entre sus publicaciones en medios especializados se encuentran: *A search for the physical content of Lüders Rule* (1990), *La objetividad del azar en un mundo determinista* (1990), *El azar objetivo como medida matemática de desorden* (1990), *Lüders Rule as a description of individual state transformations* (1991).

José Manuel Méndez Rodríguez (Reinosa, España, 1955), catedrático de lógica de la Facultad de Filosofía de la Universidad de Salamanca, está especializado en lógicas no-clásicas. Autor habitual en revistas y medios especializados con títulos como: *A Routley-Meyer semantics for Converse Ackermann Property* (1987); *Axioamatizing E→ and R→ with Anderson and Benalp's «strong and natural list of valid entailments»* (1987), *The compatibility of relevance and mingle* (1988), *Converse Ackermann Property and semiclassical negation* (1988).

Jesús Mosterín Heras (Bilbao, España, 1941) es especialista en lógica, filosofía de la ciencia, teoría de la información y teoría de la racionalidad, desarrolla la docencia en el Departamento de Lógica, Historia y Filosofía de la Ciencia de la Universidad de Barcelona. Entre otros títulos, ha publicado: *Lógica de Primer Orden* ([3]1983), *Conceptos y teorías en la ciencia* (1987), *Filosofía de la cultura* (1993), *Teoría de la escritura* (1993).

Raúl Orayen, especialista en lógica, filosofía de la lógica y de la matemática, es investigador del Instituto de Investigaciones Filosóficas de la Universidad Autónoma de México. Además de numerosos artículos, ha publicado *La ontología de Frege* (1971) y *Lógica, significado y ontología* (1989).

Lorenzo Peña Gonzalo (Alicante, España, 1944) es investigador en las áreas de lógica y metafísica en el Instituto de Filosofía del CSIC (Madrid). Entre sus publicaciones se encuentran: *El ente y el ser: un estudio lógico-metafísico* (1985), *Fundamentos de ontología dialéctica* (1987), *Rudimentos de lógica matemática* (1991), *Hallazgos filosóficos* (1992), *Introducción a las lógicas no-clásicas* (1993).

José Daniel Quesada Casajuana (Barcelona, España, 1947), es especialista en lógica y filosofía de la ciencia, disciplina que imparte en la Universidad Autónoma de Barcelona. Autor de *La Lógica y su Filosofía* (1985), así como de numerosos artículos y colaboraciones en obras colectivas y revistas especializadas.

José Antonio Robles García (México, D. F., México, 1938), es investigador del Instituto de Investigaciones Filosóficas de la Universidad Nacional Autónoma de México, especialista en historia de la filosofía y experto en la filosofía de los siglos XVII y XVIII. Cuenta, entre otros, con los siguientes títulos publicados: *George Berkeley: Comentarios filosóficos* y *Estudios berkeleyanos* (1990).

Margarita Vázquez Campos (La Coruña, España, 1961), es especialista en lógica temporal y teoría y dinámica de sistemas y desarrolla su labor académica en la Facultad de Filosofía de la Universidad de La Laguna. Autora, en colaboración, de los siguientes libros: *Computer-based management of complex systems* (1989) y *Conjuntos y proposiciones (Ejercicios de teoría intuitiva de conjuntos y lógica clásica de proposiciones)* (1990).